Chapter 2

Scattering of Electromagnetic Waves from a Spherical Boundary Using a Thin Phase Screen Model and Scalar Diffraction Theory

2.1 Introduction

This monograph focuses principally on the calculation of the electromagnetic field vector observed by a low Earth orbiting (LEO) spacecraft during an occultation of a Global Positioning System (GPS) satellite that is broadcasting a navigation signal. Of specific interest is the change in the observed signal amplitude and phase of the electric field that results from a sharp change across a spherical surface of a refraction-related property of the atmosphere. This would include a discontinuity in the atmospheric refractivity itself, or in some other quantity, such as the gradient of the refractivity caused, for example, by a change in scale height or the lapse rate of the temperature profile. Discontinuities such as these, and also milder dislocations, can result in multiple rays arriving concurrently at the LEO or, conversely, in shadow zones where no rays (according to geometric optics) arrive. Diffraction also is a by-product of these types of refractivity changes, but the amplitudes and frequencies of the diffraction fringes very much depend on the sharpness of the change and its type. Also, sometimes, caustics are present for certain LEO orbital positions.

The calculation of the observed electromagnetic field for these situations can be very difficult. Even with certain geometric assumptions in place, such as spherical symmetry, GPS/LEO orbital coplanarity, and a planar wave front for the incident wave, the exact wave-theoretic solution to this problem based on

Maxwell's equations is not readily tractable. Mie scattering theory is an example of a wave-theoretic approach to this spherical problem. Chapter 3 uses Mie scattering theory to study scattering from a spherical surface in an otherwise homogeneous medium. Chapter 5 uses a modified version of Mie scattering to deal with a scattering surface that is embedded in a refracting medium. Both of these chapters involve a full-wave theory approach, which results in spectral series representations of the electromagnetic field that are solutions to the Helmholtz equation.

Here in Chapter 2, however, we use a combination of geometric optics and scalar diffraction theory applied to a thin phase screen model to develop the phase and amplitude profile that would be observed by a LEO as a result of a change in a refractivity-related quantity. The thin screen becomes a proxy for the actual atmosphere. A vertical profile for the atmosphere-induced phase delay is embedded in the thin-screen model in such a way that the phase and amplitude profile observed by the LEO over time matches (according to geometric optics) the actual profile. This model in a ray-theoretic framework greatly simplifies the calculations. But, when diffraction effects are included, one must also deal with a convolution integral over the vertical profile in the thin phase screen. In wave theory, the spectral series representations of the electromagnetic field are essentially integrals over spectral number space. In the thin-screen approach, the integral representation of scalar diffraction effects is effectively over impact parameter space.

In Fig. 2-1 we see a ray with its point of tangency below the surface r_o, that is, $r_* < r_o$. This surface at $r = r_o$ marks a boundary separating the "+" and the "–" regimes. Across this surface a single discontinuity is assumed to occur (in an otherwise benign medium) in one of the parameters that define the relationship between the refractivity and the radial distance r. We will consider the effects of small discontinuities at r_o in three parameters: N_o, the refractivity at r_o; H_{p_o}, the pressure scale height; and $\gamma = T^{-1}dT / dr$, the normalized lapse rate of the temperature profile. The profile of the observed bending angle of the signal from an occulted GPS satellite will change as a result of a discontinuity at r_o in any one of these quantities.

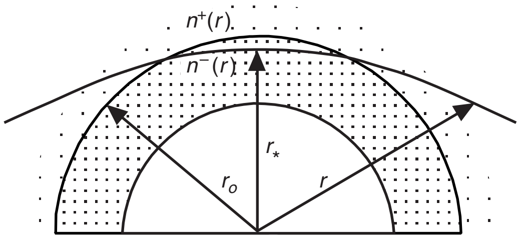

Fig. 2-1. Ray path in a stratified medium with a spherical boundary separating two regimes.

2.1.1 Multipath Scenarios

Figure 2-2 shows different multipath scenarios. Figure 2-2(a) is a schematic ray diagram for a collimated beam of plane waves encountering from the left a sphere of lower refractivity than that in the surrounding medium, a so-called

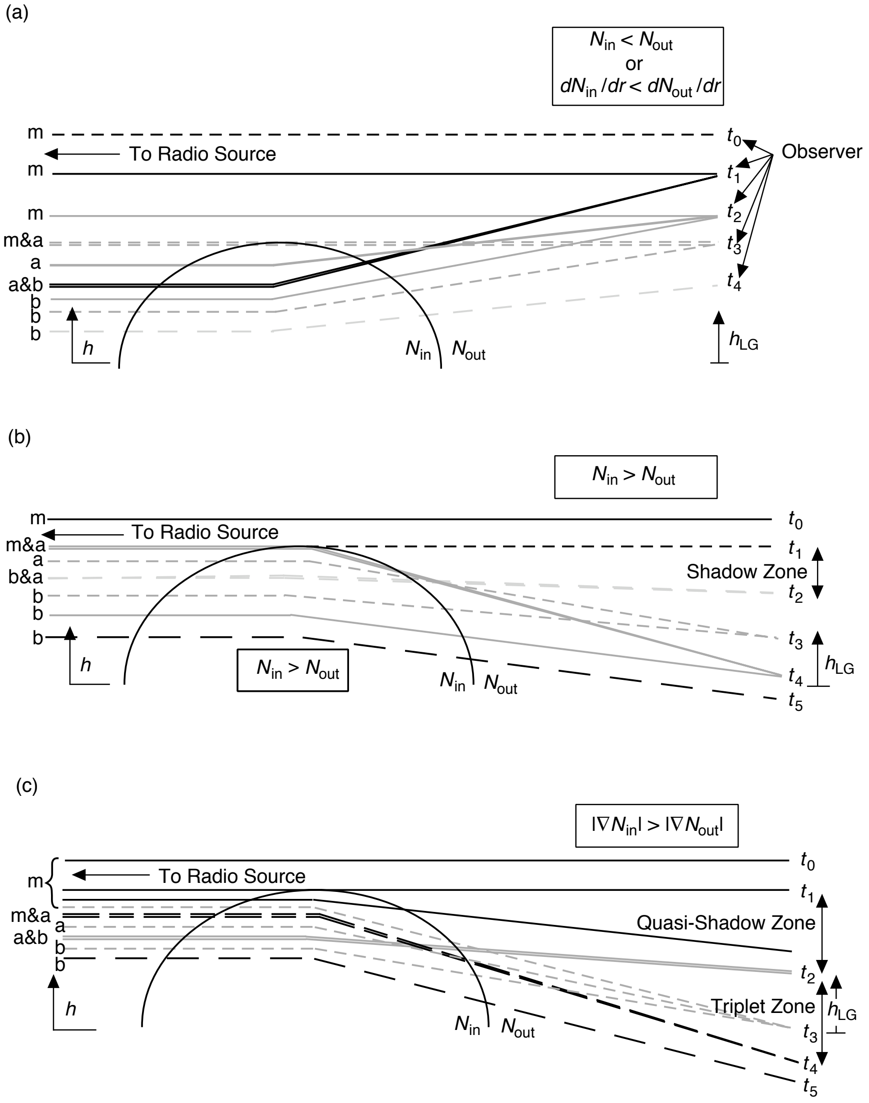
(a)
$N_{in} < N_{out}$
or
$dN_{in}/dr < dN_{out}/dr$
m
To Radio Source
m
m
m&a
a
a&b
b
b
b
h
N_{in}
N_{out}
t_0
t_1
t_2
t_3
t_4
Observer
h_{LG}
(b)
$N_{in} > N_{out}$
m
To Radio Source
m&a
a
b&a
b
b
b
h
$N_{in} > N_{out}$
N_{in}
N_{out}
t_0
t_1
Shadow Zone
t_2
t_3
t_4
t_5
h_{LG}
(c)
$|\nabla N_{in}| > |\nabla N_{out}|$
m
To Radio Source
m&a
a&b
a
b
b
h
N_{in}
N_{out}
t_0
t_1
Quasi-Shadow Zone
t_2
Triplet Zone
h_{LG}
t_3
t_4
t_5

Fig. 2-2. Multipath scenarios caused by a refracting sphere according to geometric optics. Rays arriving on the right at the same point have a common observation epoch denoted by $t_0, t_1, t_2, \cdots$: (a) $N(r)$ or its gradient is less inside than outside, (b) $N(r)$ is greater inside with super-refractivity, and (c) the gradient of $N(r)$ is greater inside without super-refractivity. The altitude of the observer is denoted by h_{LG}, and the altitude of the impact parameter of the ray is denoted by h.

sub-refraction scenario. For a setting occultation, the observer may be thought of here as traveling vertically downward on the right-hand side of the panels along the points labeled $t_0, t_1, t_2, \cdots$, which denote the epochs at which one observes the rays arriving at those corresponding points. The altitude of the observation point is denoted in the figure (relative, say, to the bottom of the frame) by the variable h_{LG}. There is a one-to-one and almost linear relation between these points in time and the values of h_{LG}, $\dot{h}_{\mathrm{LG}} \doteq D\dot{\theta}_{\mathrm{L}} = 2-3$ km/s, where D is the LEO limb distance and $\dot{\theta}_{\mathrm{L}}$ is its orbit angle rate relative to the direction of the GPS satellite and projected into the plane of propagation. The altitude of the corresponding observed ray before passing through the sphere (shown on the left edge of the figure) is denoted by h, which is closely related to the impact parameter of the ray minus the reference. For observations taken before t_1, only a single ray is encountered, and the measured phase and amplitude of the signal can unambiguously be assigned to this singlet. The epoch t_1 marks the first contact by the observer with a triplet ray system, the rays of which have been labeled in Fig. 2-2 as "m" for the main ray, "a" for the anomalous ray, and "b" for the branching ray. Rays observed at the same epoch arrive at the same time point on the right. At the position of the observer at t_1 in Fig. 2-2(a), the newly created a and b rays are collinear, but thereafter the points of tangency of the observed a and b rays, that is, their h or impact parameter values, migrate in opposite directions: for the a ray upward and for the b ray downward. Thus, at the epoch t_1, the altitude h_{LG} of the observer is stationary with respect to the altitude h of the a,b ray system. This corresponds to dh_{LG}/dh being zero at t_1 for either the a or the b ray, and having opposite polarities on these two rays for later times as they drift apart. This results in a separation in observed bending angles and/or excess Doppler, which are both offset from the bending angle and excess Doppler associated with the main ray m, the third ray. Between epochs t_1 and t_3, the observer must deal with a triplet ray system for this case. Worse scenarios with quintuplets and higher numbers can readily be constructed. As time is shown progressing in Fig. 2-2(a), the observer finally reaches the point, at epoch t_3, where the m and a rays have merged—their respective points of tangency, or h values, have met at the boundary of the sphere; dh_{LG}/dh also is zero here. This marks the end of the line for these two rays, and from this epoch onward (for example, at t_4) one has only the singlet b ray remaining; it now becomes m, the new main ray. We show later that caustic rays occur where $dh_{\mathrm{LG}}/dh = 0$, in this figure at t_1 and at t_3. The scalar diffraction version of this scenario in Fig. 2-2(a), which is discussed later, is shown in Figs. 2-10(a) and 2-10(c) and in Figs. 2-11(c) and 2-11(d).

Figure 2-2(b) shows a schematic ray diagram for the converse scenario where the refractivity is abruptly larger within the sphere than outside. This creates a super-refractivity zone. Here one observes a blackout (according to geometric optics) for the time interval from t_1 to t_2; the observer moves into a shadow zone at t_1 and exits it at t_2. This is accompanied by a flaring at t_2; at this point $dh_{\mathrm{LG}} / dh = 0$. Thereafter, the observer sees a doublet ray system. This doublet is a result of the condition $rn(r) - r_* n^-(r_*) \geq 0, \forall r \geq r_*$ being violated for tangency points nearing the boundary, a hard discontinuity in refractivity in this example, which causes the main ray to terminate abruptly at the boundary and no rays to arrive in the shadow zone. The scalar diffraction version of this scenario is discussed in Section 2.7 (see Figs. 2-11(a) and 2-11(b).

Figure 2-2(c) relaxes the discontinuity in refractivity shown in Fig. 2-2(b). It imposes a continuous refraction profile, but it invokes a large but finite gradient in refractivity between the boundary and a short distance below it. Below this level, the gradient resumes its reference profile. However, the magnitude of the gradient in the transition region is constrained to ensure that the ray existence condition $rn(r) - r_* n^-(r_*) \geq 0, \forall r \geq r_*$ holds for all values of the tangency point r_* near the boundary, thereby ensuring that rays exist for all values of the impact parameter. The radius of curvature of the rays here is greater than the spherical radius. This scenario results in a triplet ray system (for $t_2 \leq t \leq t_4$), although in Fig. 2-2(c) the m ray becomes severely defocused in this multipath region. Figures 2-10 (b) and 2-10(d) show a diffraction version of this scenario. This scenario also is further discussed in Section 2.8 in connection with an actual occultation taken by Global Positioning System/Meteorology (GPS/MET) (see Figs. 2-15 and 2-16). Evidently, on this particular occasion the tangency point of the ray(s) descended through a sharp positive radial gradient in electron density at the bottom of a sporadic E-layer.

Multipath with caustics has been seen in radio occultation observations of Uranus and Neptune [1,2]. When convective mixing of atmospheric gases plays a minor role, layering is a plausible result. To a first approximation, the multipath scenario shown in Fig. 2-2(c) probably occurred in the Uranus occultation displayed in Fig. 1-11. It shows an open-loop power spectrum over time, observed in 1986 from the Earth, of the Voyager 2 radio signal as the spacecraft followed an occulting trajectory behind Uranus [1]. The power spectrum over time is composed of contiguous power spectrum strips 10 s wide. Over roughly 1 hour the signal from Voyager passed through Uranus' atmosphere. During the 10- to 20-minute interval shown in Fig. 1-11, the signal in its descent into the atmosphere encountered a layer composed of gases of higher refractivity.

For an Earth limb sounder, one might expect to encounter this kind of feature upon first contact of the tangency point of the ray with a marine layer in the lower troposphere. However, for the Earth, the time interval for completion of the transient would be measured in a few seconds rather than the several minutes shown in Fig. 1-11 for Uranus, but the Doppler spread between tones would be comparable. We note that in Fig. 1-11, just prior to and just after the epoch of first contact with the caustic, the signal power of the main ray is very low. This suggests that it passed through a narrow transition region at the boundary with a large gradient in refractivity and, consequently, is causing severe defocusing.

A closed-loop tracking receiver is unlikely to deal adequately with the kind of multipath scenarios shown here when its signal correlation and tracking-loop feedback logic are predicated on encountering only singlet tones. Even the phase and amplitude data from a receiver that tracks in the so-called “flywheel” mode will complicate significantly the recovery process unless the trifurcation and subsequent annihilation of these tones are properly accounted for or effectively eliminated by preprocessing the data. In the flywheel mode, the receiver uses its last several reliable phase measurements to extrapolate forward in time to attempt to keep the radio frequency (RF) sample window centered in Doppler space at the most likely location of the tone(s). If the receiver had been using its last good estimate of the Doppler tone from the m ray for its flywheeling forward, what does it do at the epoch t_3, which is the end of the line? To continue onward along the evolution of the b ray with time, does one account for the integer cycle difference that has accumulated between the m and b rays during the interval from t_1 to t_3? Appendix F discusses the bias in recovered refractivity that can result from missing cycles.

2.1.2 An Overview of Chapter 2

To apply a thin phase screen model for calculating signal amplitude and phase observed by a LEO, we need a few concepts from geometric optics for the actual atmosphere, which we assume here is spherical symmetric. This includes the integral expression for refractive bending angle in terms of the radial profile for the index of refraction, the Abel transform for recovering the refractivity profile, and other related quantities. Geometric optics is used to obtain the observed change in the refractive bending angle and phase delay as a function of impact parameter due to any one of the discontinuity scenarios described earlier. Then, we introduce two thin phase screen models and discuss their use and their correspondence to the actual atmosphere. Multipath, caustics, and shadow zones are discussed in the framework of a thin-screen model, first in a geometric optics context and later in a wave-theory context. Next, to account for diffraction effects in carrier phase and amplitude, we introduce into the thin-screen model a particular scalar diffraction theory in the form of the Rayleigh–Sommerfeld integral. To numerically evaluate this diffraction integral, we introduce the stationary-phase concept and its role, not so much as a stand-alone computation technique in terms of Fresnel integrals, but as an aid for other computation techniques, and for assigning the appropriate phase profile to the thin-screen model. Numerical results for Fresnel diffraction from a discontinuous refractivity profile then are presented. Also presented are the diffraction effects from a continuous refractivity profile but with a discontinuous lapse rate.

We then use thin-screen/scalar diffraction techniques developed here to treat certain caustic, multipath, and shadow effects that can be observed by the LEO as the ray path tangency point crosses an ionosphere layer; for examples, see Figs. 2-15 and 2-16 in Section 2.8. We discuss the errors in the retrieved refractivity and temperature profile that can result from use of a straight Abel transform recovery algorithm without accounting for the deficiency in ray theory for this case. Finally, in later sections, we briefly discuss the potential of the Fresnel transform technique for enhanced resolution in thin atmosphere conditions and sketch how it might be used in multipath situations. Use of a scalar diffraction technique to sharpen the resolution of localized features is briefly covered.

2.2 Geometric Optics in a Spherical Medium

To use a thin-screen model to calculate these effects, we need some concepts from geometric optics applicable to the actual atmosphere. We need to express the refractive bending angle as a function of the index of refraction, and we need a few other related concepts.

The refractive bending angle observed by a LEO during an occultation is obtained directly from the Doppler observations made by the LEO and the precision orbit determination (POD) information about the satellites. Appendix A, see Figs. A-2 and A-3 and Eqs. (A-8) through (A-18), derives the relationship between the excess Doppler and the atmosphere-induced refractive bending angle α of a ray from the occulting GPS satellite. The excess Doppler is the observed Doppler minus the Doppler due to the relative motion between the LEO and the observed GPS satellite. Both the case where the GPS satellite is located at a finite distance from the Earth's limb and the limiting case where it is infinitely afar are given. The latter case[1] yields a particularly simple near-linear relationship between α and the excess Doppler f_D, which is given by

$$\lambda f_D = V_\perp \alpha + \mathrm{O}\left[\alpha^2\right] \tag{2.2-1}$$

---

[1] Placing the occulted GPS satellite at infinity is adequate for our purpose, but not for treating actual data. For the finite case, the bending angle equals the sum of two deflection angles, δ_G and δ_L. δ_G is the deflection angle of the ray asymptote at the GPS satellite (see Fig. A-3) relative to the straight line passing through the positions of the GPS and LEO satellites; δ_L is the deflection angle at the LEO. In practice δ_L is roughly an order of magnitude larger than δ_G. These deflection angles are mutually constrained by ray tracing or by Bouguer's law when spherical symmetry applies.

Here λ is the wavelength of the carrier of the observed signal, and $V_{\perp}$ is the component of the cross-velocity of the straight-line path between the GPS and LEO satellites lying in the propagation plane, the plane containing the GPS satellite, the LEO, and the ray. In Figs. 2-2 and 2-3, $V_{\perp} = -dh_{\mathrm{LG}} / dt$ for a setting occultation. When the LEO orbital altitude is about 10 percent of the Earth's radius, $|V_{\perp}| \cong 2.5\ \text{km/s}\ \pm 20\%$. The 20 percent spread results from the typical range of obliquity values between the propagation and orbit planes in the set of occultations used for data analysis. $V_{\perp}$ is essentially constant over the relatively short duration of an occultation. For the Earth's atmosphere at sea level, α is about 20 mrad for the dry air component; thus, the range of f_D is a few hundred hertz for dry air. Water vapor in the lower troposphere can double or triple this range.

Geometric optics as a ray theory can be developed from different approaches. One approach is to start from a wave representation based on Maxwell's equations for a harmonic wave in a refracting medium, and then take its limiting form as the wavelength of the wave is made to approach zero. The general properties of rays, such as ray path curvature as a function of refractive gradient, reflection coefficients, and so on, can be deduced. An account of this approach is given in [3]. Another approach is to start from Fermat's stationary-phase principle for rays in general and apply the Calculus of Variations to the path integral for the phase delay along the ray. Each of the rays in Fig. 2-3 satisfies Fermat's principle, which requires that each must be a path of stationary phase. In other words, the total phase accumulation or delay along the path from beginning to end would undergo only a second-order variation as

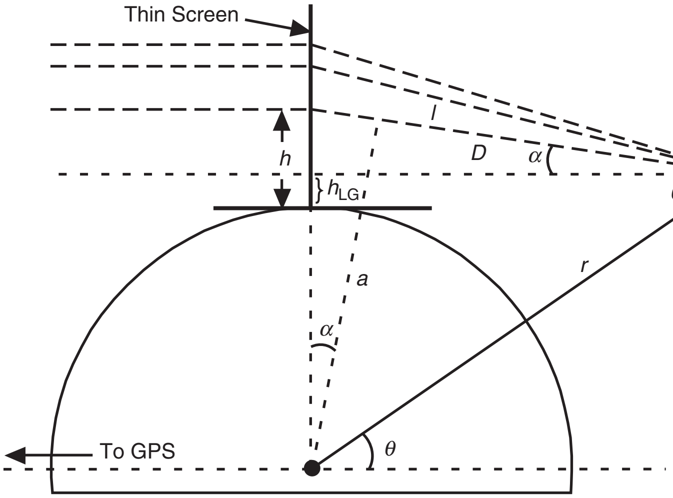

Fig. 2-3. Thin-screen geometry for a LEO. The figure shows multiple rays, with different bending angle and impact parameter values, all converging at the LEO.

a function of a nearby trial deviation of the path from the actual path, while still satisfying the boundary conditions. Usually the stationary value of the phase delay along a ray is a local minimum. However, for the scenarios shown in Fig. 2-2 the a ray, unlike the m and b rays, is a path that provides a local maximum in the phase delay. We know from the Calculus of Variations that this can happen when a ray at some interior point along its path comes in contact with the envelope associated with the family of rays whose members are generated by varying some parameter, such as the impact parameter of a ray. In optics this envelope is known as the caustic, or burning curve. Appendix A provides a brief introduction to ray theory using the stationary-phase concept. Appendix B discusses the properties of a caustic.

From Eqs. A-1 through A-5 in Appendix A, it follows that the bending angle $\alpha(r_*)$ and the impact parameter a in a locally spherical symmetric atmosphere are given by

$$\alpha(r_*) = -2a\int_{r_*}^{\infty}\frac{1}{n}\frac{dn}{dr}\frac{1}{\sqrt{n^2r^2-a^2}}dr \tag{2.2-2a}$$

$$a = n(r_*)r_* = n(r)r\sin\gamma = \text{constant} \tag{2.2-2b}$$

where γ is the angle between the radius vector and the tangent vector of the ray. When we place the GPS satellite at an infinite distance, then $\gamma = \theta + \alpha$. The quantity r_* is the radial distance of the turning point of the ray or its point of tangency with the Earth's limb (Fig. 2-1). The quantity $n(r) = 1 + N(r)$ is the index of refraction, and $N(r)$ is the refractivity. (For convenience, we define $N = n - 1$, not $N = 10^6(n-1)$; the latter is the customary definition of refractivity.) The quantity a in Eq. (2.2-2b) is the so-called impact parameter, and it is a ray-specific quantity. When spherical symmetry applies, it has a constant value when evaluated at any point (r,θ) on a specific ray. The relationship given in Eq. (2.2-2b) for a is known as Bouguer's law. This is analogous to the conservation of angular momentum in a central force field. Bouguer's law is the spherical equivalent of Snell's law.

For the spherical symmetry assumed here, and when at a specific observational epoch there is a unique ray producing the excess Doppler, then Eqs. (2.2-1) and (2.2-2) show that the Doppler observations between the LEO and occulted GPS satellite, plus the POD information about the satellites, yield both the bending angle α for the ray and its impact parameter a. Hence, one can form a data sequence $(\alpha_\kappa, a_\kappa)$, $\kappa = 1,2,\cdots,M$, from the excess Doppler sequence obtained from the M observations made over the occultation episode. This sequence $(\alpha_\kappa, a_\kappa)$ describes the evolution over time of the bending angle and impact parameter for a specific family of locally unique rays that are generated by the orbital motions of the satellites. When the geometry does not admit spherical symmetry, then the equations and procedures for recovering bending angle α and ray path tangency point r_* involve ray-tracing techniques, but they lead to the same general result, a data sequence $(\alpha_\kappa, r_{*_\kappa})$. We avoid this complication here.

When spherical symmetry does apply, and there is a unique ray from the GPS satellite to the LEO at each epoch in the observational sequence, one can invoke the Abel transform to recover the radial profile of the refractivity from the sequence $(\alpha_\kappa, a_\kappa)$. This is given by (see Appendix A)

$$N(a) \doteq \log n(a) = \frac{1}{\pi} \int_a^\infty \frac{\alpha(\xi)}{\sqrt{\xi^2 - a^2}} d\xi \tag{2.2-3}$$

Here $\alpha(\xi)$ is the functional form of the bending angle versus impact parameter ξ obtained from the data sequence $(\alpha_\kappa, a_\kappa)$, $\kappa = 1, 2, \cdots, M$. Implicit in the form for the Abel transform given in Eq. (2.2-3) is the vanishing of the bending angle for large values of the impact parameter, that is, $\alpha(\xi) \to 0$ as $\xi \to \infty$.

The gradient of $n(r)$ in Eq. (2.2-2a) is given by dn/dr. But, note that this integral readily allows a change of variable. For example, we can define another variable $\rho = kn(r)r$; Eq. (2.2-2) becomes

$$\alpha(\rho_*) = -2\rho_* \int_{\rho_*}^\infty \frac{1}{n} \frac{dn}{d\rho} \frac{1}{\sqrt{\rho^2 - \rho_*^2}} d\rho, \quad \rho_* \equiv ka = \rho \sin\gamma \tag{2.2-2$'$}$$

The functional form for the bending angle is transformed from $\alpha(r_*)$ to $\alpha(\rho_*)$, which is acceptable provided we know or can recover the functional form for $n(\rho)$, and provided no super-refractive zones exist. The integrals in Eqs. (2.2-2′) and (2.2-3) form an Abel transform pair.

A problem arises when multiple rays from the same GPS satellite meet at the LEO, which is shown in Fig. 2-3. In this case, the signal arriving at the LEO is a composite of these multiple rays. The superposition of these rays causes interference in the phase and amplitude of the signal. Spectral techniques are useful for unraveling the excess Doppler frequency and amplitude for each interfering ray. We next address multipath in a ray theoretic context.

2.2.1 Multipath, Shadow Zones, and Caustics According to Geometric Optics

A polarity change in the gradient of the bending angle is an almost sure sign that multipath and/or shadow effects will be seen, particularly if the observer is far from the location of these reversals in the gradient. Diffraction effects also are likely to be present in at least part of the observations if the change is sharp enough. Figure 2-3 provides the basic geometry for a particular thin-screen model, which we discuss later. Here h is the altitude (relative to an arbitrary but defined reference altitude R, say, sea level) of a point in a thin screen through which the ray from the occulted GPS satellite passes on its way to the LEO. The screen in this figure is oriented perpendicularly to the straight line between the LEO and the GPS satellite; h_{LG} is the altitude in the thin screen of the intersection of the LEO–GPS line. It varies nearly linearly with time during an occultation episode, about 1 minute in duration for the neutral atmosphere. Figure 2-3 shows a collimated beam from the GPS satellite to the left of the screen, which is tantamount to setting the GPS distance D_{G} from the Earth's limb to infinity. It is appropriate in this case instead to use for LEO limb distance the quantity D, which is the reduced distance between the LEO and the thin screen minus a small quantity $a\alpha$. It is given by

$$D^{-1} = D_{\mathrm{L}}^{-1} + D_{\mathrm{G}}^{-1} \tag{2.2-4}$$

where D_{L} is essentially the distance of the LEO from the limb of the Earth, and similarly for D_{G}. In calculating the phase at the LEO, use of the reduced distance D instead of D_{L} accounts for the extra phase from the wave front curvature resulting from the finite distance D_{G} of the GPS satellite from the Earth's limb. D_{G} is about 4 Earth radii, and D_{L} is about 1/2 an Earth radius; therefore, D is about 10 percent smaller than D_{L}.

From the geometry in Fig. 2-3, we see that h denotes the height at which the asymptote of the ray intersects the thin screen. From Bouguer's law in Eq. (2.2-2), we have the relationship for this height h in the thin screen:

$$\left.\begin{array}{l} h + R = a\sec\alpha = r\sin(\theta+\alpha)\sec\alpha = \\ r(\sin\theta + \cos\theta\tan\alpha) = h_{\mathrm{LG}} + R + r\cos\theta\tan\alpha \\ \text{or} \\ h = h_{\mathrm{LG}} + D\alpha(h) + \mathrm{O}\left[\alpha^2\right], \;\; D = r\cos(\theta+\alpha) \end{array}\right\} \tag{2.2-5}$$

Here the bending angle has been expressed as a function of thin-screen altitude $\alpha(h)$. We show later that when $d\left(\alpha^2\right)/da$ is less than a critical value, there is a one-to-one relationship between h and the impact parameter a. In this case, it is straightforward to transfer between $\alpha(h)$ and $\alpha(a)$.

Figure 2-3 shows an example where three altitudes result in bending angles so that the rays from these particular tangent points with the Earth's limb all simultaneously converge at the LEO. Figure 2-4 shows a hypothetical curve for h_{LG} versus h in the neighborhood of this multipath zone that might correspond to the multiple rays shown in Fig. 2-3. This depicts a classic multipath situation.

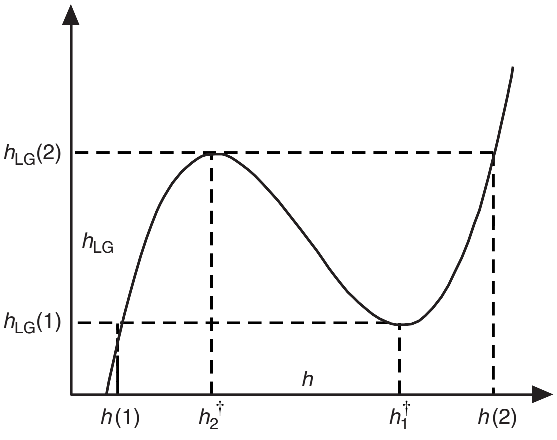

Fig. 2-4. Hypothetical curve showing thin-screen altitudes h versus h_{LG} in the neighborhood of a multipath zone. Multipath occurs within the range $h_{LG}(1) \le h_{LG} \le h_{LG}(2)$, $h(1) \le h \le h(2)$, and $h_1^\dagger$ and $h_2^\dagger$ mark the caustic contact altitudes, where $dh_{LG} / dh = 0$.

As discussed earlier in connection with Fig. 2-2, the condition $dh_{\mathrm{LG}} / dh = 0$ will be found to hold at two or more altitudes in the neighborhood defined by the turning-point altitudes for these three rays. These altitudes mark the contact points with the envelope surface to the ray family, the vertical continuum of ray paths generated by the evolution of h_{LG} over time. In the immediate neighborhood about a point $\left(h^\dagger, h_{\mathrm{LG}}^\dagger\right)$ where $dh_{\mathrm{LG}} / dh = 0$, h_{LG} has a locally quadratic (or higher) dependence on h. This results in two rays from altitudes slightly above and below $h^\dagger$ that concurrently meet at the LEO when h_{LG} is on the concave side of the h_{LG} versus h curve, and none from these two ray families when h_{LG} is on the convex side. This is depicted in Fig. 2-4 at each of the two caustic contact points. This is multipath in its simplest form.

Taking the reciprocal of dh_{LG} / dh, we have $dh / dh_{\mathrm{LG}} \to \infty$ at $\left(h^\dagger, h_{\mathrm{LG}}^\dagger\right)$; so, here there is a singularity in the density of ray paths arriving at the LEO. Caustics occur near here. Therefore, whenever the gradient of the bending angle $d\alpha / dh$, which is nominally negative for the Earth's atmosphere, takes the value $d\alpha / dh = D^{-1}$, about 0.3 mrad/km, then dh_{LG} / dh will be zero.

In Appendix A it is shown that the defocusing function for the atmosphere is given by

$$\frac{1}{\zeta} \doteq 1 - D \frac{d\alpha}{da} \tag{2.2-6}$$

The defocusing function provides a measure of the dispersal of the rays caused by the refractive gradient. Therefore, the defocusing function gives the relative amplitude of a ray at the LEO as a result of the ray having passed through the atmosphere compared to the amplitude that would have been obtained without the atmosphere. From Eq. (2.2-5) we see that in the screen $\zeta \doteq dh / dh_{\mathrm{LG}}$.

Appendix A also derives the vertical radius of the first Fresnel zone $\mathscr{F}(h)$, which is given by

$$\mathscr{F} = \sqrt{\lambda D |\zeta|} \tag{2.2-7}$$

The condition $\zeta^{-1} = 0$ marks the first contact with a caustic in geometric optics; rainbows[2] occur at such points. It also marks the first (or last) encounter of the LEO with the multiple-ray system. This condition also corresponds to the vertical diameter of the first Fresnel zone growing infinite. Therefore, geometric optics predicts an infinite amplitude at this point, but no such event is observed there, partially because of the failure of geometric optics in the neighborhood of this first contact point. The first caustic contact point corresponds to a point where not only is the phase delay along each of the multiple rays stationary with respect to a path variation, for example, the impact parameter of the path, but its second-order variation also is zero.

Geometric optics is based on a second-order stationary-phase theory. A non-zero second-order variation is required in this theory. A third-order theory is needed to handle caustics. Here, the use of the words “ray” or “path of stationary phase” implicitly assumes that a geometric optics treatment is valid. Thus, necessary conditions for validity of a geometric optics treatment are that the rays exist (see Footnote 3). Also, the altitude differences of the multipath rays at their tangency points should significantly exceed the diameter of the first Fresnel zone, and caustics are to be avoided.

A necessary condition for the existence of a caustic is that the defocusing factor $\zeta \to \infty$ at some altitude. A caustic surface will be generated by the continuum of ray paths whose points of tangency lie in some neighborhood around this critical altitude. When a caustic does occur, multipath situations also will arise where two or more ray paths arrive at the LEO from different altitudes. Appendix B gives a short discussion of caustics in a geometric optics context.

---

[2] The rainbow is a caustic phenomenon. The altitude h of the first ray making caustic contact in Figure 2-2(b) (at t_4) corresponds for a raindrop to the value of the impact parameter of the incident ray from the Sun that gives rise to a rainbow. The scattering angle of a ray, after being refracted upon entering the raindrop, internally reflected, and refracted again upon exiting, depends on the impact parameter of the ray. This scattering angle becomes stationary when the impact parameter is located about 85% (for the primary rainbow) of the raindrop radius out from the center. The impact parameters of the Sun’s rays are uniformly distributed prior to hitting the raindrop. But the exiting rays with impact parameter values in the nearby neighborhood about this 85% value “pile up” at the stationary scattering angle, about 138 deg. Therefore, the amplitudes of these rays add up constructively because they all have incurred essentially the same phase accumulation at the observer. A ”caustic” ray is the result.

Caustics also are predicted in a wave theory framework, but there their form is softened and, although brightening is predicted, the infinite amplitude predicted by (second-order) geometric optics does not appear. Caustics in a wave theory framework are discussed in Chapters 3 and 5.

2.2.2 Thin-Atmosphere Conditions

To calculate the effects of the discontinuities related to refractivity, we use certain approximations that are applicable to a thin atmosphere. By a "thin atmosphere" we mean that the atmosphere is thin geometrically and optically. Specifically, the characteristic length of the ray path through the atmosphere $\overline{L}$ defined by

$$\overline{L} = \frac{2\int_0^\infty sN[r(s)]ds}{\int_0^\infty N[r(s)]ds} \doteq 2\sqrt{\frac{2r_o H_{p_o}}{\pi(1-\beta)}} \tag{2.2-8}$$

satisfies the condition that $\overline{L} / r_o << 1$. Here s is arc length along the ray path measured from the point of tangency. The second expression on the right-hand side (RHS) of Eq. (2.2-8) applies to an atmosphere with an exponentially decreasing refractivity profile and with a scale height H_{p_o}. See Appendix A for more detail.

The parameter β is defined by

$$\beta = -\frac{r}{n}\frac{dn}{dr} \doteq \frac{rN(r)}{H_p} \tag{2.2-9}$$

where the second expression applies to an exponential model for refractivity. This parameter β is essentially the ratio of the impact parameter of a ray to its radius of curvature. It must satisfy the condition $\beta << 1$ or, equivalently, $d(rn) / dr >> 0$. In effect, the radius of curvature of the ray at any point must be much greater than the local radius of curvature of the "iso-refractivity" or equipotential surface there. Super-refractivity situations where $\beta > 1$ over substantive[3] path lengths are specifically excluded here for the reference refractivity profile. In the case of the Earth, we have the happy circumstance that for dry air $\bar{L}/r_o \approx 0.05$. At the Earth's surface the dry air component of β is about 0.2, and at the tropopause it typically is around 0.05. On the other hand, across a sharp marine layer boundary, β can exceed unity.

Equation (2.2-8) accounts for first-order ray path curvature effects through the term $(1-\beta)^{1/2}$ in the denominator. One can calculate ray path curvature effects once the refractivity profile is specified [4]. If s is arc length along the ray from its tangency point, then to first order the radial coordinate of a point on the ray is given by

$$r \doteq r_* + (1-\beta)\frac{s^2}{2r_*} \tag{2.2-10}$$

With $\beta = 0$, this gives the straight-line relationship between chord length and radial coordinate. With $\beta > 0$, the ray is bent radially inward. Both Eqs. (2.2-8) and (2.2-10) show that the thin atmosphere test fails when β nears or exceeds unity.

Atmospheric refractivity for the Earth may be classified in increasing difficulty as (1) locally spherical symmetric without multipath at the LEO, (2) locally spherical symmetric with multipath at the LEO, (3) not locally spherical symmetric because of horizontal variations, and (4) time variable through turbulence, advection, and so on. Our reference refractivity profile will satisfy the thin atmosphere conditions, and it will be spherical symmetric without multipath. We will explicitly avoid locally harsh refractivity conditions, such as critical refractivity conditions found at times in the lower troposphere, other than the discontinuities under study, which actually do include super-refractivity cases. Our discontinuities or sharp transitions are assumed to be sufficiently mild that we need not account via the Fresnel formulas for reflections from the boundary, nor for Love or Rayleigh-like waves propagating along or near the boundary surface, nor for evanescent waves. We also assume that our reference atmosphere is temporally smooth and does not cause appreciable scintillation (although scintillation is almost always present to some degree in sounding data). Even a cursory review of wave and/or ray theory literature in such disparate fields as electrodynamics and seismology will reveal that propagation across a boundary is potentially a very complicated problem without simplifying assumptions, such as those in the foregoing discussion.

---

[3] The validity of Eq. (2.2-2) requires that $rn(r) \geq a$ along the ray path. This condition is required to obtain real solutions to the Euler differential equation (see Appendix A), which all ray paths must satisfy in a spherical symmetric medium. This condition $rn(r) \geq a$ must hold at all points along the ray path to obtain a stationary-phase path satisfying the boundary conditions and passing through the turning point at r_*. In other words, if this condition is not satisfied, then there is no ray with the impact parameter value a satisfying the boundary conditions. This global condition translates into a local necessary condition (but not sufficient; see Section 6.4.2) that requires that $\beta < 1$ in some neighborhood about a turning point.

We note from Eq. (2.2-8) that $\overline{L}$ is about 400 km for dry air, but the along-track resolution δL is closer to

$$\delta L = 2\sqrt{\frac{2r_o \delta r}{1-\beta}} \approx 225\sqrt{\frac{\delta r}{1-\beta}} \text{ km} \tag{2.2-11}$$

where δr is the vertical resolution. This is the path length in a shell of thickness δr, essentially the chord length corrected to first order for ray path curvature $(1-\beta)^{-1/2}$. Using the Fresnel vertical diameter for δr, we obtain about 180 km at sea level for dry air and 280 km in the stratosphere. Because both the defocusing factor ζ and the ray path curvature parameter β depend strongly on the refractivity and its gradient, δL for the water-vapor-laden lower troposphere can vary widely. Near-super-refractivity situations where $\beta \to 1$ render the above definition of δL less meaningful.

2.3 Thin Phase Screen Models

Diffraction effects occur to a varying degree whenever a sharp change in refractivity or in one of its derivatives occurs. Sommerfeld defined the word "diffraction" more than a century ago to cover those effects not predicted by ray theory. Because diffraction effects for a spherical geometry are difficult to calculate, one often attempts to substitute a simpler model that yields basic ray theory results plus diffraction effects, and which has acceptable agreement with rigorous wave theory results. In a thin-screen model, one substitutes for the atmosphere a thin-screen proxy, nominally transparent, through which any passing ray experiences a position-dependent phase delay upon emerging from the screen. The relationship between position in a thin screen and altitude of the turning point of the actual ray in the atmosphere depends on the choice of screen and on the index of refraction profile. The thin-screen model is a surrogate for the transmission effects of the real atmosphere on the traversing GPS signal. One chooses a phase profile for the thin screen that attempts to match the LEO-observed phase delay profile. This is accomplished by assigning a phase delay, embedded in the thin screen at a given thin-screen position, that corresponds to the actual atmosphere-induced phase delay predicted by geometric optics. If a certain thin atmosphere condition holds, the assigned phase delay values in the screen can be made to be a single-valued function of the thin-screen position. In this case, the assignment leads to the proper total refractive bending angle $\alpha(a)$ when a path of stationary phase is followed from the GPS satellite along the incoming ray asymptote through the thin screen and along the outgoing ray asymptote to the LEO. When this thin atmosphere condition is satisfied, the profile of this phase delay in the thin-screen yields the LEO-observed phase and Doppler of the primary rays (excluding, for example, reflections) predicted by geometric optics for the actual atmosphere, including multipath and shadow regions where appropriate. This thin-screen model then is used to calculate the observed diffraction effects using a scalar diffraction theory such as the Rayleigh–Sommerfeld integral [5].

The thin-screen model has been discussed in many references over the years, where it has been applied to scintillation and diffraction studies [6–11]. For phase scintillation applications, [9] also discusses the wave propagation coherency conditions that an atmosphere with a finite scale height must satisfy in order for the thin-screen model to remain valid. These coherency conditions can be related in part to the “thin atmosphere” conditions given earlier in Section 2.2, namely, that β and $\bar{L}/r_o$ should be small.

Chapter 1 points out that the thin-screen concept can be extended to include multiple thin screens in tandem; see, for example, [10,11]. Here one replaces propagation through the actual atmosphere with field values holding on m thin screens, which are separated along the axis of propagation by some assigned distance D_k, $k = 1,2,\cdots, m$. The propagation of the wave between successive screens is accomplished using the Fresnel–Kirchoff scalar diffraction integral that applies asymptotically when the wavelength of the propagated wave becomes a very small fraction of the scale of the screens. The actual inhomogeneous medium in between screens is replaced by a homogeneous medium, but the phase accumulation that would have resulted is lumped into an extra position-dependent phase increment that is added to the wave at each screen. In this way the field values for amplitude and phase on one plane are mapped to the field values on the next plane, and so on. This approach has been shown to work well in a “forward” propagation process through an inhomogeneous medium.

2.3.1 The Helmholtz–Kirchoff Integral Theorem

This theorem provides the basis for scalar diffraction theory [3,12]. It may be used in either a “backward” or “forward” propagation mode, i.e., toward an emitter or away from it. In the backward mode, the scalar diffraction integral is used to map the phase and amplitude sequences measured by the LEO through the vacuum to an equivalent surface much closer to the Earth’s limb. Appendix A derives both the forward and backward modes from the Helmholtz–Kirchoff theorem. From Fig. 2-2 it is clear that, if we could move the surface containing the LEO and its vertical motion to another surface much closer to the limb, fewer rays would cross each other between the limb and the closer surface. A virtual LEO moving in the closer surface would experience reduced multipath (see Fig. 1-13). The refractivity profile of the real atmosphere is recovered from the phase and amplitude profiles mapped from the LEO trajectory to the new closer surface. The so-called “back plane” propagation method is one example of this technique [13–15]. Here the chosen surface for convenience is planar, actually a line in coplanar propagation, and it is oriented perpendicular to the LEO–GPS line, or perhaps canted slightly. In this closer “plane” with hopefully far fewer multipath episodes, one derives bending-angle and impact parameter profiles. From Eq. (2.2-1) it follows that $d\varphi_B = k\alpha ds_B$, where φ_B is the mapped phase in the back plane mounted perpendicular to the LEO–GPS line, and ds_B is an incremental arc length along the plane. Therefore, the gradient of the mapped phase along this plane yields the bending angle in that back plane, from Bouguer’s law the impact parameter. The Abel transform then yields the refractivity profile.

2.3.2 The Space Curve for Impact Parameter

There is nothing in the Helmholtz–Kirchoff integral theorem (see Appendix A), the basis for scalar diffraction theory, that requires planarity for the phase screen; it could be a curved surface. Or, if it is planar, it need not be mounted perpendicular to the LEO–GPS line; these are matters of convenience. We could as well adopt as a canting angle $\alpha_m = \alpha(a)$ for the clockwise rotation. This is suggested in [14]. In this case, the “height” h in Eq. (2.2-5) becomes $h = a - R$. It is strictly the impact parameter minus the reference radius, and it becomes an arc length along a curve defined by the intersection of the plane of propagation with the impact parameter surface. This space curve for the impact parameter in general is not straight. Since there is a one-to-one relationship between impact parameter and bending angle when spherical symmetry holds, then we are assured that $h(\alpha)$ will be monotonic. Of course, we don’t know a priori the value of $\alpha_m = \alpha(a)$ and, therefore, its shape without first processing the observations.

We show one way to recover the impact parameter curve in space coordinates from the LEO phase and amplitude observations. To simplify the geometry here we assume that the occulted GPS satellite is infinitely far in the $\theta = \pi$ direction; therefore, the angle γ in Bouguer’s law in Eq. (2.2-2) between the radius vector and the tangent vector of the ray becomes $\gamma = \theta + \alpha$. The finite case is a straightforward extension of the following, but it requires more care in handling the relative motion of the satellites during the occultation.

The phase delay φ along the ray $\mathscr{R}$ from its turning point at (r_*, θ_*) to a point (r, θ) can be obtained from its defining integral:

$$\varphi = k\int_{\mathscr{R}} nds = k\int_{r_*}^{r} n\sqrt{1+(r\theta')^2}\,dr = k\int_{r_*}^{r} \frac{n^2 r}{\sqrt{n^2r^2 - a^2}}dr \tag{2.3-1}$$

where in the last integral we have used Bouguer's law in Eq. (2.2-2) to rewrite $d\theta / dr$ in terms of the impact parameter. We assume that the point (r,θ) is out of the atmosphere where $n \equiv 1$. By integrating by parts and using Eq. (2.2-2), we obtain for the phase at (r,θ)

$$\left.\begin{aligned} &\varphi = k\left(r\cos(\theta+\alpha) + a\alpha + \int_a^\infty \alpha(\rho)d\rho \right), \\ &a = r\sin(\theta+\alpha),\ \ \alpha = \alpha(a) \end{aligned}\right\} \tag{2.3-2}$$

Setting $\alpha \equiv 0$, $\varphi \to kr\cos\theta$, which is the cumulative phase along a planar wave traveling in the $\hat{z}$ direction (see Fig. 2-5) from the line defined by $\theta = \pi / 2$. Therefore, $\varphi - kr\cos\theta$ is the excess phase caused by the refractivity. If we hold r and θ fixed, then we note from Eq. (2.3-2) that

$$\frac{\partial\varphi}{\partial a} = k\left(-a\frac{d\alpha}{da} + \alpha + a\frac{d\alpha}{da} - \alpha \right) \equiv 0 \tag{2.3-3}$$

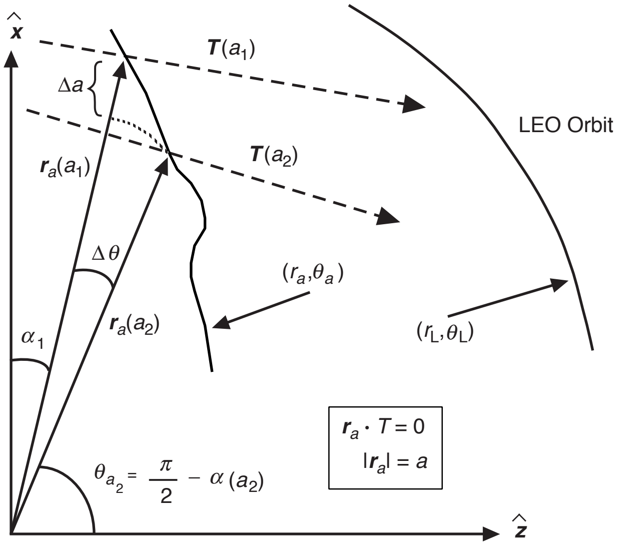
$\hat{x}$
$\boldsymbol{T}(a_1)$
Δa
LEO Orbit
$\boldsymbol{r}_a(a_1)$
$\boldsymbol{T}(a_2)$
$\Delta\theta$
(r_a,θ_a)
α_1
$\boldsymbol{r}_a(a_2)$
(r_L,θ_L)
$\boldsymbol{r}_a \cdot \boldsymbol{T} = 0$
$|\boldsymbol{r}_a| = a$
$\theta_{a_2} = \frac{\pi}{2} - \alpha(a_2)$
$\hat{z}$

Fig. 2-5. Schematic of the impact parameter space curve (r_a, θ_a), which is traced out by the tip of vector $\boldsymbol{r}_a(a)$ as a is varied. Bouguer's law requires that $|\boldsymbol{r}_a| = a$ and that $\boldsymbol{r}_a \cdot \boldsymbol{T} = 0$, where $\boldsymbol{T}$ is the unit tangent vector of the ray passing through the end point of $\boldsymbol{r}_a(a)$.

which merely underscores the stationary property of the phase on a ray, and the difficulty in using phase directly in ray theory to recover the impact parameter of the ray.

If we vary r and θ in Eq. (2.3-2), holding a fixed, we obtain

$$d\varphi = k(-a d\theta + \cos(\theta+\alpha)dr) \tag{2.3-4}$$

We now set $r = a$ and $\theta = \theta_a = \pi/2 - \alpha(a)$. By doing so, we have positioned our point (r,θ) at the tip of the impact parameter vector $\boldsymbol{r}_a(a)$. This is indicated in Fig. 2-5, which shows the geometry (but definitely not to scale) of the impact parameter space curve (r_a,θ_a) with $r_a = a$. The form for $\boldsymbol{r}_a(a)$ in terms of unit Cartesian vectors $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{z}}$ is given by

$$\boldsymbol{r}_a(a) = a(\hat{\boldsymbol{z}}\sin\alpha(a) + \hat{\boldsymbol{x}}\cos\alpha(a)) \tag{2.3-5}$$

Bouguer's law requires that $|\boldsymbol{r}_a(a)| = r_a = a$ and that $\boldsymbol{r}_a(a)$ is perpendicular to $\boldsymbol{T}(a) = \hat{\boldsymbol{z}}\cos\alpha(a) - \hat{\boldsymbol{x}}\sin\alpha(a)$, the unit tangent vector of the ray passing through the tip of $\boldsymbol{r}_a(a)$. At this point (a,θ_a), a small change in position alters the phase by an amount $d\varphi$. According to Eq. (2.3-4), to first order this is given by

$$d\varphi|_{\theta=\pi/2-\alpha} = -akd\theta + 0dr \tag{2.3-6}$$

There is no first-order variation in phase in the radial direction because the ray through the tip of $\boldsymbol{r}_a(a)$ is perpendicular to $\boldsymbol{r}_a(a)$.

This impact parameter vector $\boldsymbol{r}_a(a)$ traces out the impact parameter space curve (r_a,θ_a), as indicated by the schematic example in Fig. 2-5. The incremental arc length vector $d\boldsymbol{s}_a$ along this space curve and the angle ψ_a between $d\boldsymbol{s}_a$ and $\boldsymbol{r}_a(a)$ are given by

$$\left.\begin{aligned} d\boldsymbol{s}_a &= \frac{d\boldsymbol{r}_a}{da}da = (\hat{\boldsymbol{r}}_a - \hat{\boldsymbol{\theta}}_a a\alpha')da, \ \boldsymbol{r}_a = \hat{\boldsymbol{r}}_a a \\ \cos\psi_a &= \boldsymbol{r}_a \cdot \frac{d\boldsymbol{s}_a}{ds_a} = \frac{1}{\sqrt{1+(a\alpha')^2}}, \ \sin\psi_a = \frac{-a\alpha'}{\sqrt{1+(a\alpha')^2}} \end{aligned}\right\} \tag{2.3-7}$$

where $\alpha' = d\alpha/da$; also, $\hat{\boldsymbol{r}}_a$ and $\hat{\boldsymbol{\theta}}_a$ are unit orthogonal vectors in the rotating frame. It follows that, except at super-refractivity points, the impact parameter values are denumerated uniquely along the curve traced out by $\boldsymbol{r}_a(a)$ as a is varied. With spherical symmetry, a unique relationship holds between a and α, except at a super-refractivity point, which is discussed later.

To generate the impact parameter curve, we start from a known point on the curve where no multipath exists and, assuming that it is a high point, we continue downward. Let $\alpha(a_1,\theta_1)$, $\theta_1 = \pi/2 - \alpha(a_1)$ denote such a known point, as shown in Fig. 2-5. We assume that, in addition to knowing unambiguously the value of α at (a_1,θ_1), we also know the value of α'. The back propagation diffraction integral, for coplanar propagating in a vacuum the phase and amplitude measurements made by the LEO to another point at $\boldsymbol{r}_1$ (see Eq. (A-22) in Appendix A), is given by

$$E(\boldsymbol{r}_1) = \sqrt{\frac{i}{\lambda}} \int_C \left(\frac{E(\boldsymbol{r}_\mathrm{L})}{r_{1\mathrm{L}}^{1/2}} \exp(-ikr_{1\mathrm{L}}) \left(\hat{\boldsymbol{r}}_{1\mathrm{L}} \cdot \hat{n}(\boldsymbol{r}_\mathrm{L}) \right) \right) ds_\mathrm{L} \tag{2.3-8}$$

where $r_{1\mathrm{L}} = |\boldsymbol{r}_1 - \boldsymbol{r}_\mathrm{L}|$, $\boldsymbol{r}_\mathrm{L}$ denotes a point at the LEO, ds_L denotes an incremental arc length along the curve C defined by the trajectory of the LEO over which observations were made during the occultation, and $\hat{n}(\boldsymbol{r}_\mathrm{L})$ is the outward unit normal vector to the curve C. The intervening medium between points $\boldsymbol{r}_\mathrm{L}$ and $\boldsymbol{r}_1$ is taken as a vacuum. We apply this path integral using the LEO observations to obtain the mapped phase and amplitude at the impact parameter point $\boldsymbol{r}_1 = \boldsymbol{r}_a(a_1)$. (There are several important details required in practice related to establishing integration limits, preserving phase coherency, i.e., $|\delta r_{1L}| << \lambda$, where $\delta r_{1\mathrm{L}}$ is the error in $r_{1\mathrm{L}}$, over the span of the integral by accounting for the relative satellite motion, and maintaining phase connection. We omit those details here [13–15].) Even though this point $\boldsymbol{r}_1 = \boldsymbol{r}_a(a_1)$ is well inside the refracting atmosphere, we treat it as a proxy curve in a vacuum, just like a thin screen. Let the phase of the mapped field $E(\boldsymbol{r}_a(a_1))$ from the diffraction integral in Eq. (2.3-8) at this impact parameter point be designated as $\hat{\varphi}_1 = \hat{\varphi}(a_1)$. Next, we alter $\boldsymbol{r}_a(a_1)$ by a small amount to $\boldsymbol{r}_a(a_2)$ with $a_2 = a_1 + \Delta a$, as shown in Fig. 2-5 with $\Delta a < 0$. It follows that a first approximation for $\boldsymbol{r}_a(a_2)$ is given by

$$\boldsymbol{r}_a(a_2) \doteq \boldsymbol{r}_a(a_1) + \hat{\boldsymbol{r}}_{a_1} \Delta a - \boldsymbol{T}_1 a_1 \Delta\theta \tag{2.3-9}$$

where $\boldsymbol{T}_1 = \boldsymbol{T}(a_1)$ is the unit tangent vector in the direction of the ray passing through the point (a_1,θ_1), $\theta_1 = \pi/2 - \alpha(a_1)$. As a zeroth approximation for $\Delta\theta$, we have $\Delta\theta^{(0)} = -\alpha'(a_1)\Delta a$. To this provisional point for $\boldsymbol{r}_a(a_2)$ at $(a_1 + \Delta a, \theta_1 + \Delta\theta^{(0)})$, we use the diffraction integral again to obtain the phase $\hat{\varphi}_2^{(0)}$ from the mapped complex field $E(a_1 + \Delta a, \theta_1 + \Delta\theta^{(0)})$. Taking the difference in the mapped phases at (a_1,θ_1) and at $(a_1 + \Delta a, \theta_1 + \Delta\theta^{(0)})$,

$\Delta\hat{\varphi}^{(0)} = \hat{\varphi}_2^{(0)} - \hat{\varphi}(a_1)$, it follows from Eq. (2.3-5) that an updated estimate of $\Delta\theta$ is given by

$$\Delta\theta^{(1)} = -\frac{\Delta\hat{\varphi}_2^{(0)}}{\bar{a}k}, \quad \bar{a} = \frac{a_1 + a_2}{2} = a_1 + \frac{\Delta a}{2} \tag{2.3-10}$$

The updated estimate of θ_2 is given by $\theta_2^{(1)} = \theta_1 + \Delta\theta^{(1)}$. Mapping the LEO field again using the diffraction integral to the updated provisional point on the impact parameter curve $\left(a_2, \theta_2^{(1)}\right)$ gives us $\hat{\varphi}_2^{(1)}$. We again obtain from the difference in mapped phases $\Delta\hat{\varphi}^{(1)} = \hat{\varphi}_2^{(1)} - \hat{\varphi}_2^{(0)}$, and use Eq. (2.3-10) to obtain an updated value $\Delta\theta^{(2)}$. We iterate until convergence is achieved. This gives the point (a_2, θ_2) with $\alpha(a_2) = \pi/2 - \theta_2$. Note that at this new point (a_2, θ_2) we now have the converged phase $\hat{\varphi}(a_2) = \hat{\varphi}_2$, the bending angle $\alpha(a_2) = \alpha_2 = \pi/2 - \theta_2$, and an updated value for $\alpha_2' = -(\hat{\varphi}_2 - \hat{\varphi}_1)/a_2$ to start the next step at $a_3 = a_2 + \Delta a$. By this way of succession we generate the impact parameter curve $\boldsymbol{r}_a(a)$ over the range of impact parameter values relevant to the occultation.

The most important aspect of this approach is that it facilitates direct recovery of the refractivity profile. Since we now have recovered an unambiguous bending-angle profile $(a, \alpha(a))$, we can use the Abel transform in Eq. (2.2-2') to recover $\log n(a)$, and thence $\log n(r_*)$ from $a = r_* n(r_*)$.

If a super-refracting layer occurs at a given altitude range, then ray theory imposes restrictions on the location of tangency points in and below the layer. In a super-refracting layer located in the range $r_d \le r \le r_u$, the condition $rdn/dr + n \le 0$ holds. Here the radius of curvature of a ray would be shorter than the local radius of curvature of the equipotential surface. For tangency points in the layer and below it down to a certain critical altitude denoted by $r^c < r_d$, a ray could not escape in a spherical stratified medium. The ray with its tangency point at the lower critical altitude $r_* = r^c$ below the layer just manages to pass through the layer and escape from its upper boundary at $r = r_u$ parallel to it. This is analogous to the parallel path direction required by Snell's law of a refracted ray exiting a planar surface at the critical internal angle of incidence. It follows from Bouguer's law that the ray through this lower critical tangency point below the layer has the same impact parameter value as the ray just grazing the top of the layer where super-refractivity first sets in, i.e., $n(r^c)r^c = n(r_u)r_u$. (See Section 6.5.) These two rays have in general different bending angles. Therefore, the impact parameter space curve has a discontinuity in θ_a, i.e., two different bending angles, but with the same value for the impact parameter. Hence, $ds_a/da \to \infty$, $dx_a/dz_a \to -\tan\alpha(a)$. The break is initially parallel to the top ray just grazing the top of the super-refractivity layer, and finally it is parallel to the emerging ray from the lower critical tangency point below the bottom of the layer.

Ray theory alone cannot tell us about $n(r)$ between the upper boundary of a super-refracting layer at $r = r_u$ and the critical altitude below the lower boundary at $r = r^c$ because no ray with its tangency point within these bounds can exist in a spherical stratified medium. There is a hiatus within the bounds $r^c \le r_* \le r_u$. However, in practice a super-refractive layer tends to be relatively thin, usually caused by a marine layer. (In radio transmission research, super-refractivity is called ducting.) Therefore, one can use the recovered profiles for $n(r)$ above and below the super-refractivity zone $r^c \le r_* \le r_u$, plus some atmospheric physics, to estimate $n(r)$ in between. In wave theory, this super-refractive zone corresponds to the attenuation or tunneling part of the Airy function of the first kind. Wave theory (Section 5.8) predicts a very weak but non-zero field for this zone, analogous to the Fresnel decay in amplitude observed behind a shadow boundary from a knife-edge. Section 2.7 includes a discussion of a thin-screen/scalar diffraction model for the field observed at the LEO from a super-refractive boundary, a limiting case where $dn / dr \to -\infty$ with Δn finite. Geometric optics already becomes inaccurate near these boundaries at $r_* \le r^c$ and at $r_* \ge r_u$ (see Fig. 2-12).

The vertical resolution potential of the impact parameter space curve approach is essentially that of wave theory. This is discussed in Chapters 1 and 6, but vertical resolution is more limited by departures from spherical symmetry and other data processing properties [4,5].

Regarding caustic contact points, they have of course disappeared on the $\boldsymbol{r}_a(a)$ space curve for the impact parameter. The existence of a caustic contact point is observer-position dependent, as with rainbows. From Bouguer's law in Eq. (2.2-2), if we alter θ and take into account that a new ray with a new impact parameter value will pass through that altered point $(r, \theta + d\theta)$, we have

$$\frac{da}{d\theta} = \frac{r\cos(\theta + \alpha)}{1 - r\cos(\theta + \alpha)\alpha'} \tag{2.3-11}$$

which is zero when we set $r = a$ and $\theta = \pi / 2 - \alpha$.

On the other hand, if we set $\theta = \theta_{\mathrm{L}}$ and $r = r_{\mathrm{L}}$, then we have the defocusing-dependent relationship between a and θ_{L}, i.e., $da / d\theta_{\mathrm{L}} = \zeta_{\mathrm{L}} r_{\mathrm{L}} \cos(\theta_{\mathrm{L}} + \alpha)$, where $\zeta_{\mathrm{L}}^{-1} = 1 - r_{\mathrm{L}} \cos(\theta_{\mathrm{L}} + \alpha)\alpha'$. $da / d\theta_{\mathrm{L}}$ can be positive or negative depending on the sign of the defocusing factor ζ_{L}. If we set $\theta_{\mathrm{L}} = \dot{\theta}_{\mathrm{L}} t + \theta_{\mathrm{L}o}$, then we have the relationship between a and time, $da / dt = \zeta_{\mathrm{L}} \dot{\theta}_{\mathrm{L}} r_{\mathrm{L}} \cos(\theta_{\mathrm{L}} + \alpha)$. For anomalous rays, da / dt is positive for a setting occultation; the impact parameter migrates upward with time for these anomalous rays and at the same time downward for normal rays.

It follows that, for the purpose of forward propagating to the LEO the perturbations in phase and amplitude caused by the atmospheric refractivity perturbations, we could assign a phase profile $\varphi(a)$ to the impact parameter space curve that is given from Eq. (2.3-2) by

$$\varphi(a) = k\left(a\alpha(a) + \int_a^\infty \alpha(\rho)d\rho \right) \tag{2.3-12}$$

where $\alpha(a)$ is the reference plus perturbed bending-angle profile calculated for a spherical symmetric atmosphere from geometric optics. Moreover, given $\alpha(a)$, we can generate the impact parameter space curve $\boldsymbol{r}_a(a)$ from its definition in Eq. (2.3-6). We show in Chapter 5 using a full-spectrum wave theory (see Table 5-1) that the stationary values with respect to spectral number of the spectral density function for phase delay in wave theory, when they exist, are very closely equal to $\varphi(a)$, given in Eq. (2.3-12).

On the other hand, the impact parameter space curve generated by $\boldsymbol{r}_a(a)$ is likely to have a complicated shape in multipath situations. Computing the distance from the $\boldsymbol{r}_a(a)$ curve to the LEO is somewhat less convenient than simply using the Fresnel approximation with a planar screen. This alternate approach works well when the caustic avoidance condition is met, which relates to the thin atmosphere conditions cited above in Eqs. (2.2-8) and (2.2-9).

2.3.3 The Fresnel Phase Screen

Here we use a single planar thin-screen model to study the multipath and diffraction processes from a single boundary embedded in a laminar atmosphere possessing an otherwise smoothly varying radial gradient in refractivity. The screen is nominally oriented orthogonal to the LEO–GPS line, and the rays impinging on and emerging from the screen are the straight-line asymptotes from the LEO and the GPS satellite (see Fig. 2-3). The relationship between the altitudes h and h_{LG} is given in Eq. (2.2-5). The single thin-screen model also could be applied to multiple but spherical symmetric boundaries, given the bending-angle profile according to Snell's law that results from a ray transecting these multiple layers (see Appendix A). As an analysis technique for recovery of the refractivity profile, the usefulness of the planar thin-screen model is limited to thin atmospheres, conditions not always found, particularly in the lower troposphere. Also, if atmospheric refractivity has significant horizontal gradients, then the multiple-screen approach would be a better choice.

To maintain a one-to-one relationship between h and a, the condition $dh/da > 0$ must hold. It follows from Eq. (2.2-5) that this condition is equivalent to the condition

$$\frac{dh}{da} = \sec\alpha\left(1 + a\tan\alpha\frac{d\alpha}{da}\right) \doteq 1 + a\alpha\frac{d\alpha}{da} > 0 \tag{2.3-13}$$

which is equivalent to proscribing any caustics from occurring in the phase screen (where $dh/da = 0$). If α is given by an exponential refractivity model, then it can be shown [see Eq. (2.3-18)] that the condition in Eq. (2.3-13) is equivalent to the condition $\beta < (2\pi)^{-1/2} \approx 0.4$. This is basically double the refractive bending from dry air at sea level. In the lower troposphere, this condition can readily be violated at a marine layer boundary, but as α decreases secularly with altitude, Eq. (2.3-13) is more readily satisfied, even with larger values of $d\alpha/da$. For h lying within the range of values where the condition in Eq. (2.3-13) is violated, i.e., $dh/da = 1 + a\alpha(d\alpha/da) < 0$, there are in general at least three values of bending angle that apply. The assignment of a unique phase function to the screen [based on Eq. (2.3-7)] is not possible for this range of h values.

On the other hand, we could cant the screen clockwise by a small angle α_m. Then, in that screen the altitude relationship would be $h = a\sec(\alpha - \alpha_m) - R$, and the caustic avoidance condition becomes $1 + a(\alpha - \alpha_m)\alpha' > 0$, which might be easier to satisfy. The thin-screen phase function to be given in Eq. (2.5-1c) would have an extra factor of $\cos\alpha_m$.

2.3.4 Suitability of the Thin-Screen Model for Diffraction Analysis

The value of the thin-screen approach for qualitative study of Fresnel processes in the Earth's atmosphere partially depends on the refractivity being fairly stratified along equipotential surfaces as well as the thin screen being viewed from afar. Also, the thin atmosphere condition should hold. If the along-track inhomogeneity in refractivity is significant, then better results likely would require a multi-screen approach. We know that the fidelity of the diffraction results from the single thin-screen model is fairly good for a grazing occultation in a uniform medium from an opaque surface—e.g., the Earth's limb [16]. When $r_o >> \lambda$, the diffraction effects from a three-dimensional opaque object can be calculated by replacing the shadow zone on the surface of that object with a disk perpendicular to the incident ray path and bounded by the silhouette of the object [17]. However, a curved and nearly transparent surface, across which a discontinuity in refractivity or in one of its derivatives occurs, can support scattered rays that arise from a combination of external and internal reflections and surface propagation (evanescent waves) yielding, for example, rainbow and waveguide effects. These would be more difficult to model with the thin screen without first solving Maxwell's equations or some other wave propagation technique for the actual surface and surrounding medium and then adopting a phase and amplitude profile for the thin screen that corresponds to the actual observed phase and amplitude.

There are three-dimensional diffraction techniques discussed in the literature that use both geometric and physical optics approaches to evaluate scattering from a variety of relatively simple geometrical objects. Helmholtz's equation in a stratified medium, as a boundary value problem in Potential Theory, can formally be solved for simple surfaces, such as spheres and cylinders, and with relatively simple asymptotic boundary conditions, such as incident planar electromagnetic waves. Solutions for scattering from a conducting or a dielectric sphere in a uniform medium were developed starting about 95 years ago by G. Mie, P. Debye, G. Watson, and others. Mie scattering theory arose from the study of scattered light from droplets. Asymptotic solutions to Helmholtz's equation for $r_o >> \lambda$ also were worked out during the early decades of the 20th century, and parabolic equation techniques for wave propagation have evolved greatly in the last 20 years [18].

Chapter 5 presents a wave theory approach that deals more rigorously with the diffraction problem discussed here. There a discontinuity in a refraction-related property is embedded in a spherical atmosphere possessing a significant refractive gradient. A modified Mie scattering theory technique is developed to deal with wave propagation through the refracting medium and across the discontinuity. Good agreement holds for a single scattering surface between the thin-screen results here and the more rigorous wave-theoretic results in Chapter 5, under the caustic avoidance assumption in Eq. (2.3-13) and for positions away from external reflections and rainbow caustics. The latter are caustics arising from multiple internal reflections and refraction within the sphere.

2.3.5 A Phase Profile for the Thin Screen

For the purpose of describing diffraction and multipath processes, we use here the perpendicular mounted phase screen described above, keeping in mind the possibility that phase screen caustics might occur. Let $\varphi(h)$ be the nominal thin-screen phase delay at a thin-screen altitude h. We must now set this profile using this model so that the predicted phase observed at the LEO matches the actual atmosphere-induced phase delay observed by the LEO. Let $\psi\left(h_{\mathrm{LG}},\tilde{h}\right)$ be the cumulative phase observed by the LEO for a ray starting from the thin screen at an arbitrary altitude $\tilde{h}$. From Fig. 2-3, we see that

$$\psi = kl + \varphi = k\sqrt{\left(\tilde{h} - h_{\mathrm{LG}}\right)^2 + D^2} + \varphi(\tilde{h}) \tag{2.3-14}$$

where k is the wave number of the ray, $k = 2\pi / \lambda$. In geometric optics, a ray path is defined by the condition that the observed phase obtained from that path is stationary. The stationary-phase condition on the ray path, from the GPS satellite (infinitely afar) through the thin screen to the LEO, requires that $\partial\psi / \partial\tilde{h} = 0$; that is,

$$\frac{\partial\psi}{\partial\tilde{h}} = k\frac{\tilde{h} - h_{\mathrm{LG}}}{l} + \frac{d\varphi}{d\tilde{h}} = k\sin\alpha + \frac{d\varphi}{d\tilde{h}} = 0 \tag{2.3-15}$$

It follows for small bending angles that

$$\varphi(h) = k\int_h^\infty \sin(\alpha(h'))dh' \doteq k\int_h^\infty \alpha(h')dh' \tag{2.3-16}$$

Here it is assumed that $\alpha(h) \to 0$ as $h \to \infty$. The LEO-observed stationary phase is given by $\psi(h_{\mathrm{LG}}) = \psi[h_{\mathrm{LG}}, h(h_{\mathrm{LG}})]$, where h is the altitude in the thin screen providing the stationary-phase value for the LEO located at h_{LG}, and it is given in terms of the impact parameter by Eq. (2.2-5). It may not be unique if multipath is present. Comparing the phase $\varphi(h)$ in Eq. (2.3-16) for the planar thin screen with the phase $\varphi(r,\theta)$ given in Eq. (2.3-2) for a ray at the point (r,θ), the difference is the term $r\cos(\theta + \alpha) + a\alpha$. Through second order, this term is just $r\cos\theta\sec\alpha$, the slant distance to the point h in the screen, which from Eq. (2.3-14) is $\left(\left(\tilde{h} - h_{\mathrm{LG}}\right)^2 + D^2\right)^{1/2}$.

2.3.6 Bending-Angle Perturbations

To apply the appropriate phase profile $\varphi(h)$ for the thin screen according to Eq. (2.3-16), we need the appropriate form for $\alpha(r_*)$. It is convenient to break $\alpha(r_*)$ into two parts: the reference bending-angle profile $\alpha(r_*)$ from a reference refractivity profile, and the perturbed component of the bending angle $\delta\alpha(r_*)$ due to the refractivity-related discontinuity. For the reference profile, we use the bending angle from an exponential profile for the atmospheric refractivity. This is given by

$$n = 1 + N_o \exp\left(-\frac{r - r_o}{H_{p_o}}\right) \tag{2.3-17}$$

Appendix A [Eq. (A-30)] derives the form for $\alpha(r_*)$ that this refractivity profile generates. It is given by

$$\alpha(r_*) = \sqrt{\frac{2\pi r_*}{H_{p*}}} N(r_*)\left(1 + (\sqrt{2} - 1)\beta_* + 0.28\beta_*^2 + \cdots\right) \tag{2.3-18}$$

Here $N(r) = n(r) - 1$ and $\beta = |n'r/n| = Nr/H_p$. This expression is very accurate for thin atmospheres, i.e., $\beta <\approx 1/2$.

We now need expressions for $\delta\alpha(r_*)$ for the different cases. A ray with its point of tangency below the discontinuity travels through both the "+" and the "–" regimes when its radius of curvature is greater than r_o, as indicated in Fig. 2-1. It follows that the bending angle observed by the LEO is given from Eq. (2.2-2) by

$$\alpha(r_*) = \\ \left\{ \begin{array}{l} -2a^- \left(\displaystyle\int_{r_*}^{r_o} \left(\frac{n'}{n}\right)^- \frac{dr}{\sqrt{(n^- r)^2 - (a^-)^2}} + \int_{r_o}^{\infty} \left(\frac{n'}{n}\right)^+ \frac{dr}{\sqrt{(n^+ r)^2 - (a^-)^2}} \right), \ r_* \le r_o \\ \alpha^+(r_*), \ r_* > r_o \end{array} \right\} \tag{2.3-19}$$

where $a^- = r_* n^-(r_*)$, and where $n^+(r)$ and $n^-(r)$ are functions describing the index of refraction in the + and – regimes, respectively. Here $n' = dn/dr$. In Eq. (2.3-10), $\alpha^+(r_*)$ is the bending angle for points of tangency of the ray lying in the + regime where $n^+(r)$ applies, and, therefore, it is given by Eq. (2.2-2) with $n = n^+(r)$ or by Eq. (2.3-18) when the refractivity profile is given by Eq. (2.3-17). Equations (2.2-2) and (2.3-19) are valid only over integration intervals where $n(r)$ is differentiable. These equations have to be appended with discrete Snell's law terms at points where $n(r)$ is discontinuous, and when $n^-(r_o) > n^+(r_o)$, the inequality $r_* \le r_o$ in Eq. (2.3-19) also must be amended to account for critical refraction. For the case where the gradient of $n(r)$ is at least piecewise continuous, Eq. (2.3-19) can be recast into the form

$$\alpha(r_*) - \alpha^-(r_*) = \\ \left\{ \begin{array}{l} -2a^- \displaystyle\int_{r_o}^{\infty} \left[\left(\frac{n'}{n}\right)^+ \frac{1}{\sqrt{(n^+ r)^2 - (a^-)^2}} - \left(\frac{n'}{n}\right)^- \frac{1}{\sqrt{(n^- r)^2 - (a^-)^2}} \right] dr, \ r_* \le r_o \\ \alpha^+(r_*) - \alpha^-(r_*), \ r_* > r_o \end{array} \right\} \tag{2.3-20}$$

Here, $\alpha^{-}(r_*)$ has a connotation similar to $\alpha^{+}(r_*)$; i.e., it is the bending angle that would be obtained from Eq. (2.2-2) if $n^{-}(r)$ applied to the entire atmosphere. Equation (2.3-20) may be generalized from a single boundary geometry to the multiple-layer case (see Appendix A).

The evaluation of the integral in Eq. (2.3-20) requires differential techniques, which will be found in Appendix A. Equation (2.3-20) involves terms of the form $\alpha^{-}(r_*)-\alpha^{+}(r_*)$ plus additional terms to account for the differences between a^{-}, a^{+}, and $a_o^{\pm}=r_o n^{\pm}(r_o)$. For example, if we write $\alpha^{\pm}(r_*)=\alpha\left[p^{\pm},r_*\right]$, where p is the parameter to be varied, then we can linearize $\alpha^{+}(r_*)-\alpha^{-}(r_*)$ by the form

$$\alpha^{+}(r_*)-\alpha^{-}(r_*)\doteq\frac{\partial\alpha}{\partial p}\Delta p \tag{2.3-21}$$

It is shown in Appendix A, for example, that when p is the scale height,

$$\frac{\partial\alpha}{\partial H_{p_o}}\hat{=}\frac{\alpha}{H_{p_o}}K_H=\frac{\alpha}{H_{p_o}}\left(-\frac{1}{2}\ +\ \mathrm{O}\left[\beta,\left(\gamma H_{p_o}\right)\right]\right) \tag{2.3-22}$$

Also, when p is the normalized lapse rate $\gamma=\left(dT/dr\right)T^{-1}$, which is a constant, one obtains

$$\frac{\partial\alpha}{\partial\gamma}\hat{=}\alpha H_{p_o}K_\gamma=\alpha H_{p_o}\left(\frac{3}{8}\ +\ \mathrm{O}\left[\beta,\left(\gamma H_{p_o}\right)\right]\right) \tag{2.3-23}$$

In Appendix A, the integrals in Eq. (2.3-20) are computed to zeroth order in β (i.e., no ray path curvature corrections are applied, but for Case A, β is included), and to first order in the discontinuous parameter of interest for the three cases cited above: discontinuous refractivity, discontinuous scale height, and discontinuous lapse rate. The reference refractivity and refractive bending-angle profiles are given for a constant lapse rate by Eqs. (A-37) through (A-39).

2.3.7 Case A: A Discontinuity in Refractivity

Case A may be considered as a limiting case for a ray that crosses a very sharp boundary, such as a well-delineated marine layer in the lower troposphere or a sporadic E-layer in the ionosphere. Let $n(r)$ have a discontinuity across the surface $r=r_o$. Then $n(r)$ has the form

$$n(r)=n^{-}(r)+\mathscr{H}\left(r-r_o\right)\left[n^{+}(r)-n^{-}(r)\right] \tag{2.3-24}$$

where $\mathscr{H}(r-r_o)$ is the Heaviside function and $n^+(r)$ and $n^-(r)$ are well-behaved reference functions, as obtained, for example, from Eqs. (A-37) and (A-38). Therefore, the gradient of $n(r)$ has the form

$$\frac{dn}{dr}=\frac{dn^-}{dr}+\mathscr{H}(r-r_o)\left[\frac{dn^+}{dr}-\frac{dn^-}{dr}\right]+\delta(r-r_o)\left[n^+(r_o)-n^-(r_o)\right] \tag{2.3-25}$$

where $\delta(r-r_o)$ is the Dirac delta function. Let σ be defined by

$$\sigma^2=\frac{(r_o-r)}{H_{p_o}},\quad r\le r_o \tag{2.3-26}$$

where H_{p_o} is the pressure scale height at $r=r_o$. It can be shown (Appendix A) that an expression for Eq. (2.3-20) valid to first order in ΔN but without ray path curvature corrections is given by

$$\begin{aligned}\alpha(r_*)-\alpha^-(r_*)&\doteq\alpha^-(r_o)\left[\left(1+\frac{\Delta N_o}{N_o}\right)F(\sigma_+)-F(\sigma)\right]\\&\quad+2\left(\sin^{-1}\left(\frac{r_*n_*^-}{r_on_o^+}\right)-\sin^{-1}\left(\frac{r_*}{r_o}\right)\right),\quad r_*\le\tilde{r}\le r_o,\\&\alpha^+(r_*)-\alpha^-(r_*),\quad r_*>r_o\end{aligned} \tag{2.3-27}$$

where $\Delta N_o=n^+(r_o)-n^-(r_o)$ and where $F(x)$ is given by

$$F(x)=(1-\operatorname{erf}(x))\exp(x^2) \tag{2.3-28}$$

The quantity σ_+ is given by

$$\sigma_+^2=\sigma^2+\left(\frac{\Delta N_o}{N_o}\right)\left(\frac{N_or_o}{H_{p_o}}\right) \tag{2.3-29}$$

which is essentially equal to σ, except near $\sigma=0$. The second term in Eq. (2.3-29) is just the fractional change in refractivity times the ratio of r_o to the radius of curvature of the ray, both of which are nominally small quantities. When $\Delta N_o<0$, $\sigma_+=0$ corresponds to the critical internal reflection condition and $\tilde{r}$, the maximum value of r_* at this critical point, is given by $\tilde{r}=r_on^+(r_o)/n^-(r_o)\doteq r_o(1+\Delta N_o)$. In geometric optics, no rays exist from the GPS satellite to the LEO with turning points located in the interval $\tilde{r}<r_*\le r_o$.

One can improve the accuracy of Eq. (2.3-27) by including bending along the ray path of integration in Eq. (2.3-20); a power series in β would result. Equation (A-49) in Appendix A accounts for most of the first-order bending for Case A.

Figure 2-6(a) shows $\alpha(r)-\alpha^{-}(r)$ near $r=r_o$ based on a numerical integration of Eq. (2.2-2a) plus the Snell term. The exponential refractivity profile given in Eq. (2.3-17) is used for the reference. The values of the reference parameters used for this figure are $N_o=200\times10^{-6}$, $r_o-R_E\approx\ 3\,\text{km}$, $H_{p_o}=7\,\text{km}$, $\gamma=0$, and $\Delta N_o/N_o=-1/20$. Most of the jump in α arises from the Snell term: $\Delta\alpha\doteq2\left(-2\Delta N_o\right)^{1/2}$.

2.3.8 Case B: A Discontinuity in Scale Height

Let ΔH_{p_o} denote a discontinuity in a piecewise constant pressure scale height, which occurs across the surface $r=r_o$. In this case, the gradient of $n(r)$ is discontinuous across this surface, but $n(r)$ is differentiable at interior points of the integration intervals defined by Eq. (2.3-20). From Appendix A, Eqs. (A-44) and (A-45), it can be shown that to first order in ΔH_{p_o}

$$\alpha(r_*)-\alpha^{-}(r_*)\doteq$$

$$\left\{\begin{array}{l}\dfrac{\Delta H_{p_o}}{H_{p_o}}\alpha^{-}\left(r_o\right)\left[\left(K_H-\sigma^2\right)F(\sigma)+\dfrac{\sigma}{\sqrt{\pi}}\right]+\mathrm{O}\left[\left(\dfrac{\Delta H_{p_o}}{H_{p_o}}\right)^2\right],\ r_*\le r_o\\[2ex] \alpha^{+}(r_*)-\alpha^{-}(r_*),\ \ r_*\ge r_o\end{array}\right\} \tag{2.3-30}$$

where K_H is given by Eq. (2.3-22) for $\gamma=0$, and $\sigma^2=\left(r_o-r_*\right)/H_{p_o}$. Appendix A also discusses the $\gamma\neq0$case. Depending on the physical circumstances present at a boundary layer in the real atmosphere, one could use a linear combination of Eqs. (2.3-27) and (2.3-30) to model the effect of a discontinuity in temperature or a boundary demarcating a lower regime containing water vapor. The values of the parameters used in Fig. 2-6(b) are $\alpha^{-}\left(r_o\right)=5.44$ mrad, $\Delta H_p/H_p=-1/7$, $r_o-R_E=10$ km, $H_{p_o}=7$ km, and $\gamma=0$.

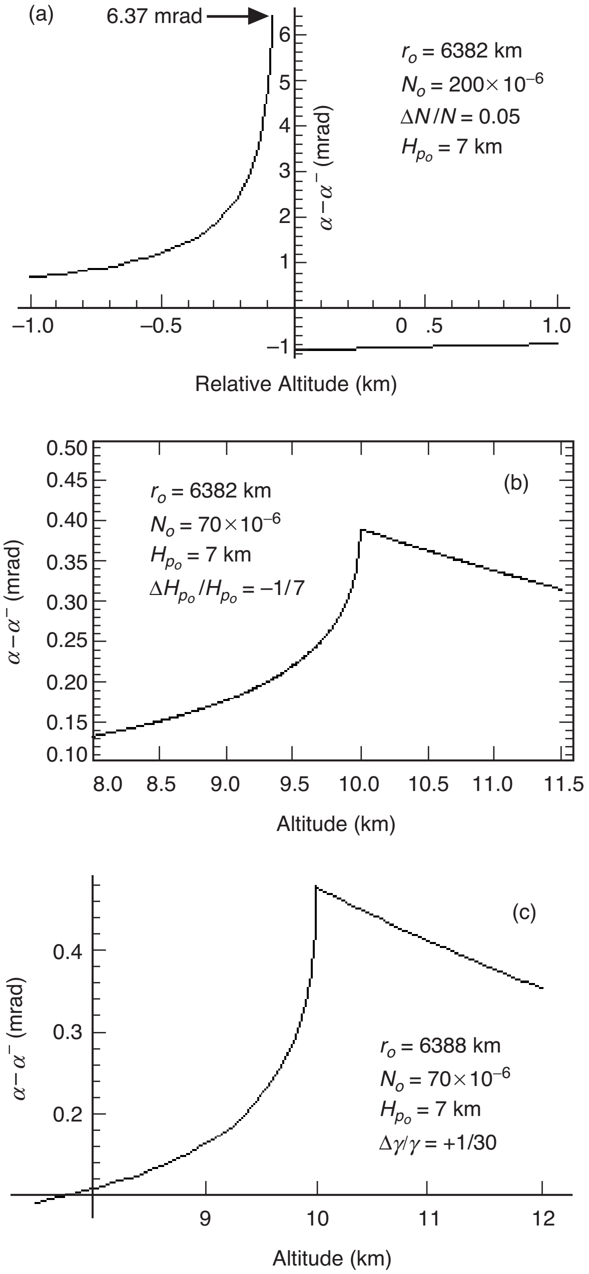

Fig. 2-6. Perturbations in bending angle profile for different refractivity scenarios: (a) perturbation in bending angle for Case A; no rays exist in the super-refractivity region where r_o – 79 m < r_* < r_o; (b) perturbation for Case B; and (c) perturbation for Case C. All figures use an exponential reference refractivity profile.

2.3.9 Case C: A Discontinuity in Lapse Rate

Let Δγ denote a discontinuity in the piecewise constant normalized lapse rate across the surface $r = r_o$. In this case, the gradient of $n(r)$ is discontinuous across this surface, but $n(r)$ is differentiable at all interior points of the integration. From Appendix A, Eqs. (A-44) and (A-45), it follows that Eq. (2.3-20) becomes to first order in $\left(\Delta\gamma H_{p_o}\right)$

$$\alpha(r_*) - \alpha^-(r_*) \doteq$$

$$\left\{\begin{array}{l} \left(\Delta\gamma H_{p_o}\right)\alpha^-\left(r_o\right)\left[\left(K_\gamma + \dfrac{3\sigma^2}{2} + \dfrac{\sigma^4}{2}\right)F(\sigma) - \dfrac{1}{\sqrt{\pi}}\left(\dfrac{5\sigma}{4} + \dfrac{\sigma^3}{2}\right)\right],\ r_* \le r_o \\ \alpha^+(r_*) - \alpha^-(r_*),\ r_* \le r_o, \end{array}\right\} \tag{2.3-31}$$

where K_γ is given by Eq. (2.3-23) (or by Eq. (A-40) for the $\gamma \ne 0$ case), and $\sigma^2 = \left(r_o - r_*\right)/H_{p_o}$. Figure 2-6(c) shows the behavior of $\alpha(r) - \alpha^-(r)$ near $r = r_o$. The values of the reference parameters used to obtain this figure are $N_o = 70 \times 10^{-6}$, $r_o - R_E = 10$ km, $H_{p_o} = 7$ km, $\gamma = 0$, and $\Delta\gamma = +1/30$, or about a 7 K/km discontinuity in lapse rate.

We note in all three cases that $\alpha(r_*) = \alpha^+(r_*)$ when $r_* > r_o$; also, for Cases B and C, $\alpha(r_*)$ is continuous across the surface $r = r_o$. It follows for Cases B and C that the difference in altitudes of the points of tangency of the two rays about the surface $r = r_o$, which is given by

$$\left\{r_*^+ - r_*^- \doteq D\left[\alpha^+\left(r_o^+\right) - \alpha\left(r_o^-\right)\right],\quad D^{-1} = D_{\text{LEO}}^{-1} + D_{\text{GPS}}^{-1}\right\} \tag{2.3-32}$$

approaches zero. We will see below that when a caustic surface is generated, one has to be more careful about the definition of the condition $r_* \to r_o$.

2.4 Multipath Using a Thin Phase Screen Model

We show now some examples of the effect of these discontinuity surfaces on the relationship between the altitude of the ray path turning point and the altitude of the LEO–GPS line. Figures 2-7(a) through 2-7(d) show the relationship between h and h_{LG}, based on the thin-screen relationship given in Eq. (2.2-5), in the vicinity of the discontinuity at r_o for Cases A, B, and C. The same parameter values hold here that were used in Fig. 2-6.

Figure 2–7(a), which corresponds to the scenario shown in Fig. 2–2(b), shows a super-refractivity situation. This results from a hard transition in

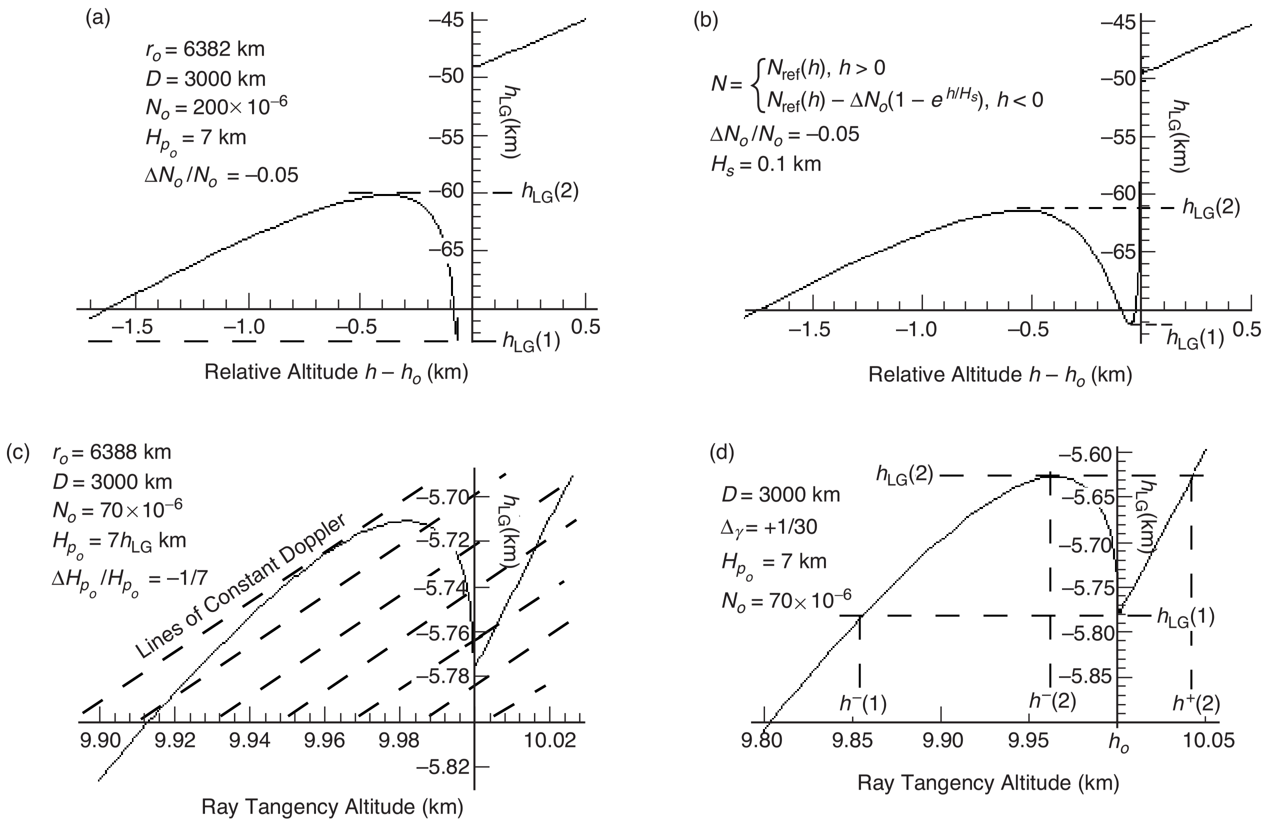

Fig. 2-7. Thin-screen relationship between altitude h of the impact parameter and altitude h_{LG} of the LEO–GPS straight line for different refractivity perturbation scenarios: (a) Case A, a super-refractivity scenario with $N_o^- > N_o^+$, (b) the same as (a) except that the profile of N has been softened to the profile given in the panel, (c) Case B, a scale height discontinuity, and (d) Case C, a lapse rate discontinuity.

refractivity (infinite gradient) at the boundary, a Case A scenario. Here a multipath doublet configuration results. The doublet begins at an altitude in h-space of about 79 m below the boundary (corresponding to the critical reflection point $\tilde{r}$) and continues down to an altitude $h(1)$, or about 1.6 km below the boundary. No rays exist for tangency points lying between the boundary and 79 m below. Figure 2–7(a) also shows the shadow zone in h_{LG}-space ($\sim\!-60.5 \le h_{\mathrm{LG}} \le \sim\!-49.3$ km or, equivalently, 4 to 5 s of elapsed time) within which no signal at all (according to geometric optics) will be received by the LEO. Below the first contact point with the caustic in h_{LG}-space, the altitude difference between the doublet rays grows to well in excess of the Fresnel diameter.

Figure 2-7(b) uses the same conditions that were used in Fig. 2-7(a) except that the discontinuity in refractivity has been replaced by a sharp continuous change over a narrow transition zone just below the boundary. Although the magnitude of the refractivity gradient is large in this zone, it is bounded so that the ray existence condition $\beta < 1$ holds for all values of the impact parameter. By softening the discontinuity, one obtains a triplet ray system and two caustics. This yields a quasi-shadow zone where the highly defocused main ray continues, and also a triplet ray system between the contact points with the two caustics, one at $\left(h(1), h_{\mathrm{LG}}(1)\right)$ and the other at $\left(h(2), h_{\mathrm{LG}}(2)\right)$. The caustic contact points mark the upper and lower boundaries of the multipath zone.

Figure 2-7(c) shows for Case B the range of ray path altitudes and h_{LG} values over which three (when $\Delta H_p < 0$) mutually interfering ray paths will be received by the LEO. Contours of constant excess Doppler also are overlaid on this figure. The excess Doppler value for the dashed contour through the point at $h_{\mathrm{LG}} = -5.78$, $h = 10$ km, is about 80 Hz. The spacing between contours is about 100 mHz.

Figure 2-7(d) shows for Case C the range of ray path altitudes and h_{LG} values over which three mutually interfering ray paths will be received by the LEO when $\Delta\gamma > 0$. The point where $dh_{\mathrm{LG}} / dh = 0$, at a ray path altitude of $h = h^{-}(2) \cong 9.96\,\mathrm{km}$ in Fig. 2-7(d), is a singularity point where the vertical radius of the first Fresnel zone approaches infinity, which is the condition for the existence of a caustic. For Case C, we see from Fig. 2-7(d) that the altitude differences between the tangency points of the three mutually interfering ray paths depend non-linearly on the location of h_{LG} within the multipath zone $h_{\mathrm{LG}}(1) \le h_{\mathrm{LG}} \le h_{\mathrm{LG}}(2)$. These multiple rays arise as a result of an abrupt downward increase in refractivity or decrease (in magnitude) of its gradient. Figure 2-7(d) shows that the maximum separation, $h_o - h^{-}(1)$, is about 170 m for a temperature lapse rate discontinuity of +7 K/km at $h_o = 10$ km, which is small compared to the local mean Fresnel zone but still a significant source of interference. The half-width in Doppler space of the local first Fresnel zone is given by

$$f_{\mathscr{F}} = \dot{h}_{LG} \frac{|\zeta(h)|}{\mathscr{F}(h)} \tag{2.4-1}$$

where $\dot{h}_{\mathrm{LG}}$ is about $2.5\ \mathrm{km/s} \pm 20\%$. For the assumed ambient medium at an altitude of 10 km, $f_{\mathscr{F}}$ is roughly 2 Hz. The maximum separation in ray path tangency points in Fig. 2-7(d), which is $(h(1) - h_o)$, corresponds to a tonal separation of about 1 Hz, well within the half-width of the first Fresnel zone and well within the 25–Hz bandwidth of the high sample rate mode in a TurboRogue GPS flight receiver. Figures 2–7(a) through 2-7(d) also may be used to obtain the time and Doppler intervals between caustic events. The abscissa should be multiplied by approximately 5 to yield a Doppler scale in hertz, and the vertical scale should be divided by approximately 2.5 to obtain a time scale in seconds. Appendix C provides further details on the separation geometry at the tangency points of the rays. It also shows the near-quadratic dependence of the separation scale (and, therefore, the difference in frequency of the tones in Doppler space) on ΔH_p for Case B and $\Delta\gamma$ for Case C. Caustics produce inherently nonlinear effects, and their existence depends on the polarities of $\Delta\gamma$ and ΔH_p.

Because the Doppler spreads in Figs. 2–7(b) and 2-7(c) are small compared to the Doppler spread of the first Fresnel zone, use of geometric optics to predict received signal amplitude and phase would be inaccurate. A diffraction treatment should be used.

When $\Delta\gamma < 0$, the left-hand portion of the curve in Fig. 2-7(c) that is less than $h = h_o = 10$ km will be flipped about the horizontal axis (at $h_{\mathrm{LG}} = h_{\mathrm{LG}}(1) = -5.78$); no caustic occurs and no multipath arises due to this kind of discontinuity. However, severe defocusing will occur at the boundary; a quasi-shadow zone results when the refractivity gradient is steep. Appendix C provides further details and an example of this converse case.

For Case B, no caustic surface is generated when $\Delta H_p / H_p > 0$ at $h = h_o$, but shadowing may be severe if the gradient is large. Diffraction effects will soften the shadowing.

We also can estimate the relative intensity (in a geometric optics context) and the relative phases of the multipath rays observed by the LEO. To obtain the relative intensity, we use the defocusing function $\zeta(h)$, which is given by Eq. (2.2–6). For Case C, for example, $|\zeta(h)|^{1/2}$ is shown in Figs. 2–8(a) and 2–8 (b) for $\Delta\gamma < 0$ and $\Delta\gamma > 0$, respectively. It should be noted, however, that the altitude separation of the multipath rays for this value of $\Delta\gamma$ is well within the first Fresnel diameter, and, therefore, the geometric optics approach will break down. Here $\zeta(h)$ will not provide a realistic measure of relative intensities; a wave theory approach is required. Figure 2-8(b) exhibits the flaring of intensity based on $\zeta(h)$ in the vicinity of the singularity point where $dh_{\mathrm{LG}} / dh \rightarrow 0$, which marks the onset of a caustic condition. Here $|\zeta|^{1/2} \rightarrow \infty$ in the vicinity of the caustic contact point. The scalar diffraction versions of these cases are given in Figs. 2-10(c) and 2-10(d). Irrespective of whether or not the altitudes of the tangency points of the multiple rays lie within the Fresnel zone, a ray optics approach based on stationary-phase theory to only second order breaks down at the point of contact with a caustic. A third-order stationary-phase theory is given in Appendix D.

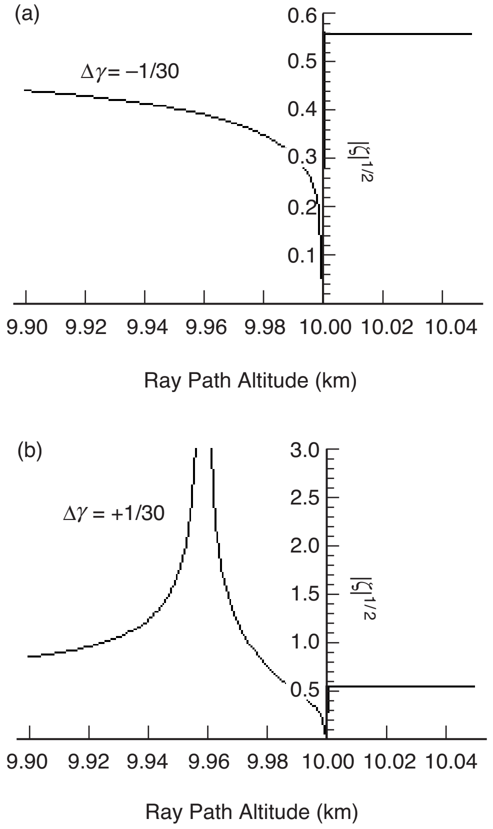

Fig. 2-8. Defocusing function $|\zeta|^{1/2}$ for Case C: (a) $\Delta\gamma < 0$; a shadow zone results; and (b) $\Delta\gamma > 0$; $\zeta \rightarrow \infty$ near the caustic contact. The scalar diffraction versions are shown in Figs. 2-10(c) and 2-10(d).

The phase difference between multipath rays observed by the LEO is given by the difference in values of the Fresnel phase function [see Eq. (2.5-1)] for the two rays. We show later that this is given by

$$\Delta = \Phi^- - \Phi^+ =$$

$$\frac{2\pi}{\lambda D}\left(\frac{1}{2\zeta^+\left(r_*^+\right)}(r_*^- - r_*^+)^2 + D\int_{r_*^-}^{\infty}\left(\alpha^- - \alpha^+\right)dr_* \right) + \mathrm{O}\left[\frac{\left(r_*^- - r_*^+\right)^3}{\lambda D^2} \right] \tag{2.4-2}$$

where $r_* - R$ is the altitude of a ray at its point of tangency with the Earth's limb. The first term on the RHS of Eq. (2.4-2) results from the increase in path length due to the additional bending $\delta\alpha$. The second term results from the additional delay through the atmosphere. For cases with a caustic, the observed phase Φ^- must be further delineated, for example, by Φ^a and Φ^b and r_*^a and r_*^b, to indicate the multipath branch to which one is referring. Again, geometric optics breaks down for the small values of $\Delta\gamma$ used in these examples and also near the point of first contact with the caustic. The actual observed phase differences will differ significantly from those predicted in Eq. (2.4-2) because of diffraction effects.

Finally, one can carry out the vector addition of the multipath signals

$$E\exp(i\Psi) = E^+\exp\left(i\Psi^+\right) + E^a\exp\left(i\Psi^a\right) + E^b\exp\left(i\Psi^b\right) \tag{2.4-3}$$

using Eqs. (2.2-6) and (2.4-2) for each ray to obtain a total electric field vector observed by the LEO as predicted by geometric optics. When the altitude differences of the tangency points of the multiple signals are significantly larger than $2\mathscr{F}(h)$ and well away from caustics, this approach usually should be valid.

2.5 Scalar Diffraction: The Rayleigh–Sommerfeld Integral

To calculate wave theory effects from these perturbations in refractivity on bending angle, we use scalar diffraction theory applied to a thin-screen phase profile $\varphi(h)$, which is defined in Eq. (2.3-16) in terms of the bending-angle profile. Scalar diffraction theory is derived from the Helmholtz–Kirchoff integral theorem from classical electrodynamics to relate the amplitude and phase distributions of an electromagnetic wave over a surface to the amplitude and phase at a point interior to the surface [3,12]. This theorem is valid when the wavelength of the wave is small compared to the scale of the radiating surface. Appendix A presents a version of this surface integral; also, its contraction to a two-dimensional coplanar geometry is derived there. It provides the basis for scalar diffraction theory. As was pointed out in Section 2.3, Eq. (2.3-8), one can use this integral to map the observed amplitude and phase measurements made by the LEO backwards to an equivalent set of observations on a surface closer to the Earth's limb. This is the basis for the back-plane propagation technique used in radio occultations to reduce multipath [13–15].

Similarly, the Helmholtz–Kirchoff surface integral also can be used to map the phase and amplitude forward from an emitting surface to the LEO [4,5]. The forward one-dimensional version of this integral [see Appendix A, Eq. (A-22)] leads directly to the Rayleigh–Sommerfeld integral for scalar diffraction. We adopt a thin-screen model to provide the phase and amplitude of the emitting source in the screen. For the purpose of forward propagating to the LEO, we use a simple phase screen mounted perpendicular to the LEO–GPS line. This should work well provided that the perturbation in refractivity does not violate the uniqueness condition $dh/da>0$ or, equivalently, $1+a\tan\alpha(d\alpha/da)>0$. This ensures a one-to-one relationship between h and a. We also use in this integral the Fresnel approximation. Here the ray path length from a height h in the screen to the LEO is given by $l \doteq D+\left(h-h_{\mathrm{LG}}\right)^2/2D$, (Fig. 2-3). For typical LEO orbits, $D>>|h-h_{\mathrm{LG}}|$, so this approximation is adequate. For a LEO orbit altitude of 600 km, D is roughly 3000 km, whereas the range of $h-h_{\mathrm{LG}}$ is only a few tens of kilometers.

Let $E\left(h_{\mathrm{LG}}\right)$ and $\psi\left(h_{\mathrm{LG}}\right)$ be the normalized amplitude and phase that are observed by the LEO at an altitude h_{LG} as a result of the Earth's intervening atmosphere. If the Earth's atmosphere were absent, the amplitude of the signal at the LEO would be unity and the phase would be kD. Diffraction effects from the Earth's opaque limb are ignored here. It can be shown (see Appendix A) that $E\left(h_{\mathrm{LG}}\right)$ and $\psi\left(h_{\mathrm{LG}}\right)$ are given through the diffraction integral formulation by

$$E\left(h_{\mathrm{LG}}\right)\exp\left(i\psi\left(h_{\mathrm{LG}}\right)\right)=\frac{1}{1+i}\exp(ikD)\sqrt{\frac{2}{\lambda D}}\int_{-\infty}^{\infty}A(h)e^{i\Phi\left(h,h_{\mathrm{LG}}\right)}dh, \quad (2.5\text{-}1a)$$

$$\Phi\left(h,h_{\mathrm{LG}}\right)\doteq\varphi(h)+\frac{\pi}{\lambda D}\left(h-h_{\mathrm{LG}}\right)^2, \quad (2.5\text{-}1b)$$

$$\varphi(h)=k\int_{h}^{\infty}\alpha(h')dh' \quad (2.5\text{-}1c)$$

We define $\Phi\left(h,h_{\mathrm{LG}}\right)$ to be the Fresnel phase function for the screen. Equation (2.5-1b) shows that it equals the thin-screen phase function $\varphi(h)$ plus the extra geometric delay from the bending angle, $k(l-D)$. For a transparent phase screen, $A \equiv 1$. In actuality, the integration limits in this integral are finite, but one can for convenience set these limits to infinity. If one is interested in edge phenomena, such as knife-edge diffraction effects from the Earth's limb, one can set $A = 0$ below a certain altitude. We will show using the stationary-phase technique that, when the stationary-phase points are located well away from questionable boundaries, the resulting error from using infinite integration limits is negligible because the significant contributions to the integral come only from neighborhoods where the phase is stationary, or nearly so.

As a check, let us set $\varphi(h) \equiv 0$ and $A \equiv 1$. Then the Fresnel phase function becomes simply the extra geometric phase delay from any point in the screen to the LEO, $\Phi \doteq \pi\left(h - h_{\mathrm{LG}}\right)^2 / \lambda D$. The integral in Eq. (2.5-1) reduces to a complete complex Fresnel integral with a value of $(1+i)(\lambda D/2)^{1/2}$. Thus, $E\left(h_{\mathrm{LG}}\right) = 1$ and $\psi\left(h_{\mathrm{LG}}\right) = kD$, as predicted.

There are alternative methods for evaluating the convolution integral in Eq. (2.5-1). One approach much favored prior to the advent of modern computers was the saddle-point method. Here we discuss the stationary-phase technique, which expresses the results in terms of Fresnel integrals. However, the stationary-phase technique as a computational technique is limited to a few ideal situations with isolated stationary-phase points. When complexities of a real atmosphere are introduced, this approach becomes cumbersome. Even handling caustics becomes a problem unless a higher-order Taylor series expansion is used in Eq. (2.5-1). The technique does provide some insights, however, and also asymptotic forms that are useful in determining the limiting value of a numerical integration. It is in that spirit that the discussion in the following section is presented.

2.6 The Stationary-Phase Technique

The stationary-phase technique uses the fact that the principal contributions to an integral with a rapidly oscillating kernel come from those neighborhoods where it has the least change. Traditionally the technique mainly has been applied to neighborhoods around well-isolated stationary-phase points or near end points associated with a boundary of some sort. In this technique, the integral in Eq. (2.5-1a) is evaluated in terms of a Fresnel integral by expanding $\Phi\left(h, h_{\mathrm{LG}}\right)$ in a Taylor series in h about the thin-screen altitude $h = h^*(h_{\mathrm{LG}})$ providing the least change in $\Phi\left(h, h_{\mathrm{LG}}\right)$ with h, while retaining only terms up to the second degree. This technique fails in the case of multiple stationary points when they are not sufficiently isolated to ensure adequate phase windup in the neighborhood between them. It also fails when the expansion of the Fresnel phase to only the quadratic term is insufficient, for example, when the quadratic term is zero or near zero, which occurs near a caustic. A third-order treatment is needed for caustics.

We show an example where the technique is sound. In this case, we assume that there is a single stationary-phase point, well away from any boundaries that the geometry might include. If we expand the Fresnel phase $\Phi(h,h_{\mathrm{LG}})$ in Eq. (2.5-1a) in a Taylor series, we obtain

$$\left.\begin{aligned}
&\Phi(h,h_{\mathrm{LG}})=\Phi(h^*,h_{\mathrm{LG}})+\left.\frac{\partial\Phi}{\partial h}\right|_{h^*}(h-h^*)+\frac{1}{2}\left.\frac{\partial^2\Phi}{\partial h^2}\right|_{h^*}(h-h^*)^2+\cdots,\\
&\Phi(h^*,h_{\mathrm{LG}})=k\int_{h^*}^{\infty}\alpha(h')dh'+\frac{\pi}{\lambda D}(h^*-h_{\mathrm{LG}})^2,\\
&\left.\frac{\partial\Phi}{\partial h}\right|_{h^*}=k\left(\frac{1}{D}(h-h_{\mathrm{LG}})-\alpha(h)\right)^*,\quad \left.\frac{\partial^2\Phi}{\partial h^2}\right|_{h^*}=\frac{\pi}{\lambda D}\left(1-D\frac{d\alpha}{dh}\right)^*
\end{aligned}\right\}\tag{2.6-1}$$

Now we set $h^*=h_{\mathrm{LG}}+D\alpha(h^*)$ to null the first-order term. This condition is identical with the thin-screen relationship in Eq. (2.2-5). Then the diffraction integral in Eq. (2.5-1) becomes

$$\left.\begin{aligned}
&E(h_{\mathrm{LG}})e^{i\psi(h_{\mathrm{LG}})}=e^{i\left(kD+\Phi(h^*,h_{\mathrm{LG}})\right)}\frac{1}{1+i}\sqrt{\frac{2}{\lambda D}}\int_{-\infty}^{\infty}e^{\left(i\frac{\pi}{\lambda D\zeta^*}(h-h^*)^2\right)}dh,\\
&\zeta^{-1}=\left(1-D\frac{d\alpha}{dh}\right)^*
\end{aligned}\right\}\tag{2.6-2}$$

Here ζ is the defocusing function [Eq. (2.2-6)] evaluated at the stationary-phase altitude h^*. We assume that it is positive in this example, but on anomalous rays $\zeta<0$. Now we make a change of integration variable $v=(2/\lambda D\zeta)^{1/2}(h-h^*)$ to obtain

$$\left.\begin{aligned}
&E(h_{\mathrm{LG}})e^{i\psi(h_{LG})}\doteq\sqrt{\zeta}e^{i\left(kD+\Phi(h^*,h_{\mathrm{LG}})\right)}\frac{1}{1+i}\int_{-\infty}^{\infty}e^{i\frac{\pi}{2}v^2}dv=\sqrt{\zeta}e^{i\left(kD+\Phi(h^*,h_{\mathrm{LG}})\right)},\\
&kD+\Phi(h^*,h_{\mathrm{LG}})=kD\left(1+\frac{1}{2}\alpha^2(h^*)\right)+k\int_{h^*}^{\infty}\alpha(h)dh,\\
&h^*=h_{\mathrm{LG}}+D\alpha(h^*)
\end{aligned}\right\}\tag{2.6-3}$$

Implicit in the integration here is that the infinite limits on the integral are valid. Otherwise, an incomplete Fresnel integral would result, yielding diffraction fringes. Moreover, we have assumed here that h^*, the stationary value for h, is unique. In the Fresnel phase expression, the term $D\left(1+\alpha^2/2\right) \doteq D\sec\alpha$ is the slant range from the LEO to the stationary-phase point in the thin screen. The second term, the integral term, is the total refractivity-induced phase delay, that is, it is the extra phase delay incurred by the ray after passing completely through the refracting atmosphere. We note the defocusing term $\zeta^{1/2}$ in the amplitude term in Eq. (2.6-3), in accordance with geometric optics.

Neither of the integration assumptions used to obtain Eq. (2.6-3) is likely to hold when refractive boundaries have to be considered, such as one of the scenarios described by Cases A, B, or C. We need to develop the perturbation to the nominal thin-screen phase-delay profile that results from a discontinuity on the surface $r = r_o$. For a ray descending through the + regime of the atmosphere, the nominal phase delay in the thin-screen model is taken to be that corresponding to the bending-angle profile $\alpha^+(r)$, which is given by Eq. (2.3-18) (see also Appendix A). Because the turning point of this ray is above the boundary, it is unfettered by the effects of the discontinuity below. The thin-screen phase function may be broken into two parts:

$$\varphi(h) = \left\{ \begin{array}{l} \varphi_{\text{Ref}}(h), \ \ h \ge h_o, \\ \varphi_{\text{Ref}}(h) + \delta\varphi(h), \ \ h \le h_o \end{array} \right\} \tag{2.6-4}$$

where $\varphi_{\text{Ref}}(h)$ is the reference thin-screen phase function derived from the integral in Eq. (2.3-16) in terms of the reference form for the bending angle for $\alpha^+(r_*)$, nominally that given in Eq. (2.3-18). Similarly, the perturbed thin-screen phase function $\delta\varphi(h)$ is obtained from Eq. (2.3-16) and the form for the perturbed bending angle, for example, Eq. (2.3-31) for Case C. Thus,

$$\delta\varphi(h) = \left\{ \begin{array}{l} 0, \ \ h \ge h_o \\ k\displaystyle\int_h^{h_o} \delta\alpha(h')dh', \ \ h \le h_o \end{array} \right\} \tag{2.6-5}$$

Here $h = n(r_*)r_* \sec\alpha - R$. The quantity $\delta\alpha(h) = \alpha(r_*) - \alpha^+(r_*)$ is the perturbation in refractive bending angle. Thus, there is a break in the diffraction integral in Eq. (2.5-1). We can use any one of Eqs. (2.3-27), (2.3-30), or (2.3-31) combined with the appropriate expression for $\alpha^-(r_*) - \alpha^+(r_*)$ to develop explicit expressions for $\delta\alpha(r)\ \delta\varphi(h)$ and for $\delta\varphi(h)$, which then can be used with $\varphi_{\text{Ref}}(h)$ in the diffraction integral to calculate the LEO-observed phase and amplitude perturbations. Appendix A gives explicit forms for $\delta\varphi(h)$ for the different discontinuity scenarios discussed here. Here h denotes a height in the thin screen, but r_* denotes a geocentric radial position in the atmosphere of the ray path tangency point with the Earth's limb. Also, $r_* - R$ denotes the height in the atmosphere of the tangency point; h is essentially equal to $r_* n(r_*)\sec\alpha - R$, as has already been discussed.

As the point of tangency of the ray traversing the atmosphere approaches the boundary, a perturbation in phase is observed by the LEO because of the discontinuity. This results from changes in both the atmospheric delay and the refractive bending angle. Figures 2–9(a) through 2–9(c) show the Fresnel phase function $\Phi(h, h_{\mathrm{LG}})$ versus h and the incomplete diffraction integral $\left|\int_h^\infty \exp\left[i\Phi(h', h_{\mathrm{LG}})\right]dh'\right| / (\lambda D)^{1/2}$ for three different h_{LG} values. A Case C scenario has been used. These figures should be compared to the thin-screen Case C scenario shown in Fig. 2–7(d), except that there the discontinuity in lapse rate is 1/7 the magnitude used to generate Figs. 2–9(a) through 2-9(c). The integral is obtained from a numerical integration. The phasor $\exp\left[i\Phi(h, h_{\mathrm{LG}})\right]$ in this integral rapidly oscillates except near stationary-phase points. Therefore, care must be exercised in evaluating the limiting value of the integral. Figures 2-9(a) and 2-9(b) show two cases where h_{LG} lies outside of the multipath zone; in Fig. 2-9(a) h_{LG} is well above the zone, and in Fig. 2-9(b) it is just below. In Fig. 2-9(c), h_{LG} lies in the multipah zone, that is, where $h_{\mathrm{LG}}(1) < h_{\mathrm{LG}} < h_{\mathrm{LG}}(2)$. Here $h_{\mathrm{LG}}(1)$ and $h_{\mathrm{LG}}(2)$ mark the lower and upper boundaries of the multipath zone, respectively. From Fig. 2-7(d), it follows that the lower limit, $h_{\mathrm{LG}}(1)$, is independent of the magnitude of the discontinuity, such as $\Delta\gamma$, and depends only on the altitude of the discontinuity and the nominal value of $\alpha^+(r_o)$. However, $h_{\mathrm{LG}}(2)$ corresponds to a nominal ray path altitude *above* h_o (when $\Delta\gamma > 0$); that is, it is the altitude of tangency point for the main ray [the m ray in Fig. 2–2(a)]. It depends on the magnitude of the discontinuity.

When h_{LG} does not lie within a multipath zone, as shown in Figs. 2-9(a) and 2-9(b), and the convexity of the Fresnel phase function $\Phi(h, h_{\mathrm{LG}})$ is such that it rapidly increases as h moves away from its stationary-phase value, then the stationary-phase approximation for the diffraction integral may be accurate. This would hold for Fig. 2-9(a), even though there are reversals in polarity of the convexity of $\Phi(h, h_{\mathrm{LG}})$ (around $h = 10\,\mathrm{km}$), because $|\partial\Phi/\partial h| >> 0$ in that neighborhood. There is a unique and isolated stationary-phase point.

However, when the convexity of $\Phi(h, h_{\mathrm{LG}})$ not only reverses sign but also induces a significant slowdown in the growth rate of $\Phi(h, h_{\mathrm{LG}})$ to near zero, which is shown in Fig. 2-9(b) (around $h = 10\,\mathrm{km}$), care in using the stationary-phase approximation must be exercised. This figure typifies the case where a quasi-stationary-phase altitude makes a small but significant contribution to the overall integral in Eq. (2.5-1), which affects both the amplitude and phase of the observed signal. A geometric optics approach to situations like this would be compromised *even though there is a unique stationary-phase altitude* for the example in Fig. 2-9(b) (at about 7.4 km).

Finally, Fig. 2-9(c) shows $\Phi(h, h_{\mathrm{LG}})$ when h_{LG} falls within the multipath zone between the two caustic contact points, which results in a triplet of

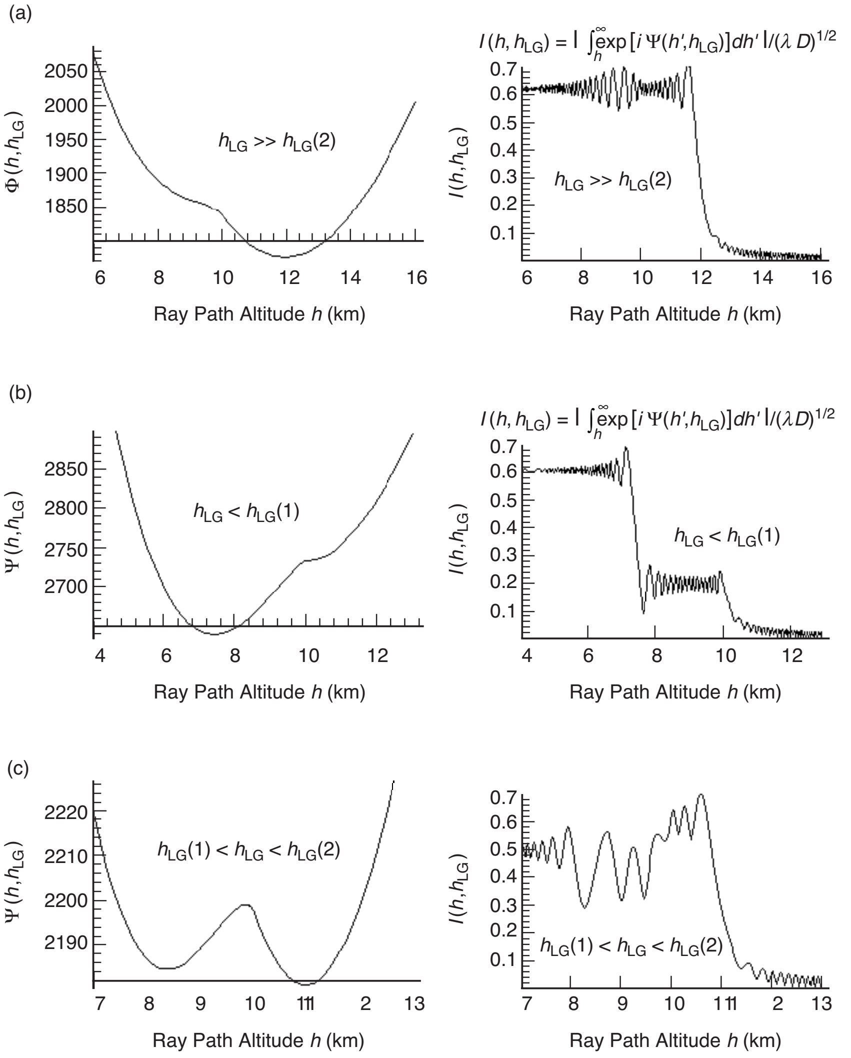

Fig. 2-9. Pairs of figures showing Fresnel phase Φ (left) and the incomplete scalar diffraction integral $I(h, h_{\mathrm{LG}})$ (right) for three different fixed altitudes of the LEO–GPS line h_{LG}. A Case C discontinuity in lapse rate at 10 km has been used with $\Delta\gamma = +7/30$. In (a), h_{LG} lies well above the upper boundary of the multipath zone at $h_{\mathrm{LG}}(2)$, (b) h_{LG} lies below the lower boundary at $h_{\mathrm{LG}}(1)$ but near it, and (c) h_{LG} lies in the multipath zone.

stationary-phase points. Here the stationary-phase approximation is likely to be seriously compromised. Multipath rays with narrow separations at their tangency points comparable to or smaller than the radius of the first Fresnel zone are difficult for the stationary-phase approximation. This is because the technique in these circumstances tends to yield a biased phase error in the individual E vector computation for each stationary-phase point, i.e., each ray, which can significantly alter the resultant E vector after vector addition of the contributions from the multiple ray segments. Near a caustic contact point at $\left(h^{\dagger}, h_{\mathrm{LG}}^{\dagger}\right)$, $\Phi\left(h, h_{\mathrm{LG}}\right)$ must be expanded in a power series in $\left(h-h^{\dagger}\right)$ that includes third-order terms to obtain sufficient accuracy.

Similarly, regarding Fig. 2-9, geometric optics should be accurate for situations depicted by Fig. 2-9(a), less accurate for Fig. 2-9(b), and inadequate for Fig. 2-9(c) because of co-mingling of the stationary-phase contributions from overlapping neighborhoods in impact parameter.

One can generalize Fig. 2-9(c) to situations where $\Phi\left(h, h_{\mathrm{LG}}\right)$ supports multiple multipath zones. It should be clear that for n multipath zones one generally will have up to $2n + 1$ stationary-phase points, i.e., $2n + 1$ rays, and also at least $n + 1$ caustics. Whenever a situation arises where the stationary-phase technique degrades in accuracy, so also will geometric optics as an approximate second-order description of electromagnetic wave processes. It should be pointed out, however, that in multipath situations where the separation altitudes between the tangency points of the rays are large compared to their respective first Fresnel zones, then a geometric optics approach can be sound. However, in the vector addition of these multiple rays to compute the complete field vector, their relative phases must be maintained with high accuracy.

2.6.1 Necessary Conditions for Validity of the Geometric Optics Approach

Let h^a, h^b, and h^c denote successive ray path altitudes at which the Fresnel phase has stationary values for a given value of the LEO–GPS straight line altitude above the Earth's limb, h_{LG} (see Fig. 2-3). Then necessary conditions for ensuring accuracy of the stationary-phase approximation technique with multiple stationary-phase points would be

$$\left.\left|\Phi\left(h^{a}, h_{\mathrm{LG}}\right)() - \Phi\left(h^{b}, h_{\mathrm{LG}}\right)\right| >> \pi, \quad \left|\Phi\left(h^{b}, h_{\mathrm{LG}}\right) - \Phi\left(h^{c}, h_{\mathrm{LG}}\right)\right| >> \pi \right\} \tag{2.6-6}$$

These conditions are tantamount to requiring the altitudes of the stationary-phase points to be separated by a distance much greater than the first Fresnel zone. However, for cases with a single stationary-phase point, as demonstrated by Fig. 2-9(b), or even for multiple stationary-phase points where the conditions in Eq. (2.5-8) hold, we also must ensure that $\Phi(h, h_{\mathrm{LG}})$ has no near-stationary points elsewhere. A strong condition would require that there be no polarity changes in the convexity of $\Phi(h, h_{\mathrm{LG}})$ over the h–space. A somewhat relaxed condition would require that $|\partial\Phi/\partial h| >> 0$ within any neighborhood bounded by reversals in polarity of the convexity of $\Phi(h, h_{\mathrm{LG}})$. Appendix D, which is principally concerned with third-order stationary-phase theory to deal with caustics, provides a more explicit bound on $|\partial\Phi/\partial h|$ for zero convexity points that are isolated ($|\Delta\Phi| >> \pi$) from end points and stationary-phase points.

2.7 Numerical Results Using Thin-Screen/Scalar Diffraction

Figures 2–10 and 2–11 show for several discontinuities, both positive and negative, the resulting phase and amplitude perturbations, $\psi(h_{\mathrm{LG}})$ and $E(h_{\mathrm{LG}})$. We refer to these as “Fresnel effects,” but in fact they are a combination of both diffraction and interference. Away from caustic rays and deep shadow zones, the interference in amplitude and phase is largely predictable by complex addition of the rays from geometric optics. These figures are based on a numerical integration of Eq. (2.5-1). Figure 2–10 shows a Case C scenario, a lapse rate discontinuity. Figure 2–11 shows a Case A scenario, a discontinuity in N. In both the figures, the exponential refractivity model given in Eq. (2.3-17) was used to generate the reference phase profile for the Fresnel phase function given in Eq. (2.5-1) for the thin screen. The perturbed Fresnel phase was obtained from Eq. (2.6-5) using the perturbed bending-angle profile given in Eq. (2.3-27) for Case A and Eq. (2.3-31) for Case C. To simplify the calculations, the atmospheric altitude difference $r_* - r_o$ was replaced directly by the thin-screen altitude difference $h - h_o$ without converting to the impact parameter value.

These numerical integrations were aided by the stationary-phase technique, which was used to isolate the neighborhoods in h-space that contribute to the integral. Numerical integration is aided by using the asymptotic forms for the incomplete Fresnel integrals, which these diffraction integrals assume in their limit. These asymptotic forms also provide the characteristic wavelengths and magnitudes of the asymptotic modulations in signal phase and amplitude. The asymptotic expansions for the Fresnel integrals can be written in the form

$$\int_0^x \exp\left(i\frac{\pi}{2}y^2\right)dy = \frac{1+i}{2} - \left(g(x) + if(x)\right)\exp\left(i\frac{\pi}{2}x^2\right) \qquad (2.7\text{-}1)$$

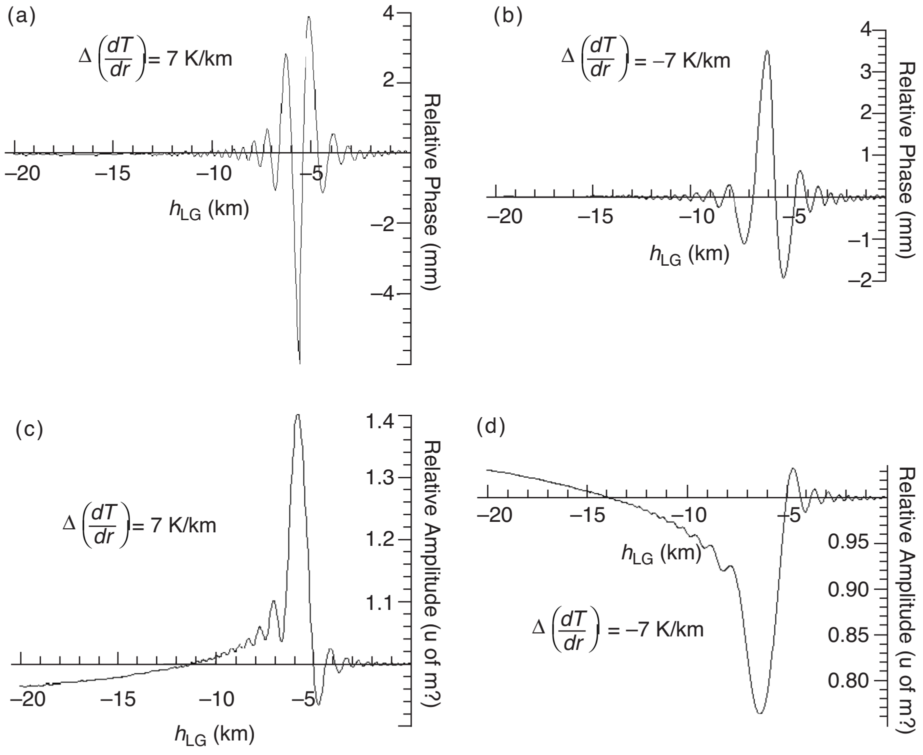

Fig. 2-10. Fresnel perturbations in received signal due to a discontinuity $\Delta\gamma = \pm 1/30$ in lapse rate at an altitude of 10 km. The phase variations are shown in millimeters for (a) $\Delta\gamma = 1/30 > 0$ and (b) $\Delta\gamma < 0$. The geometric optics terms for the reference and perturbed phase have been removed to stop phase windup. The amplitude variations are shown for (c) $\Delta\gamma = 1/30 > 0$ and (d) $\Delta\gamma < 0$.

where $f(x)$ and $g(x)$ are given by

$$\left. f(x) = \frac{1}{\pi x} - \frac{3}{\pi^3 x^5} + \cdots, \quad g(x) = \frac{1}{\pi^2 x^3} - \frac{5}{\pi^4 x^7} + \cdots \right\} \tag{2.7-2}$$

By fitting the free parameters (scale factor and mean) associated with these functions in Eq. (2.7-1) to the numerical integration of the diffraction integral in its asymptotic regime, one can calculate the limiting value of the integral without actually numerical integrating to the limit, thereby reducing computations.

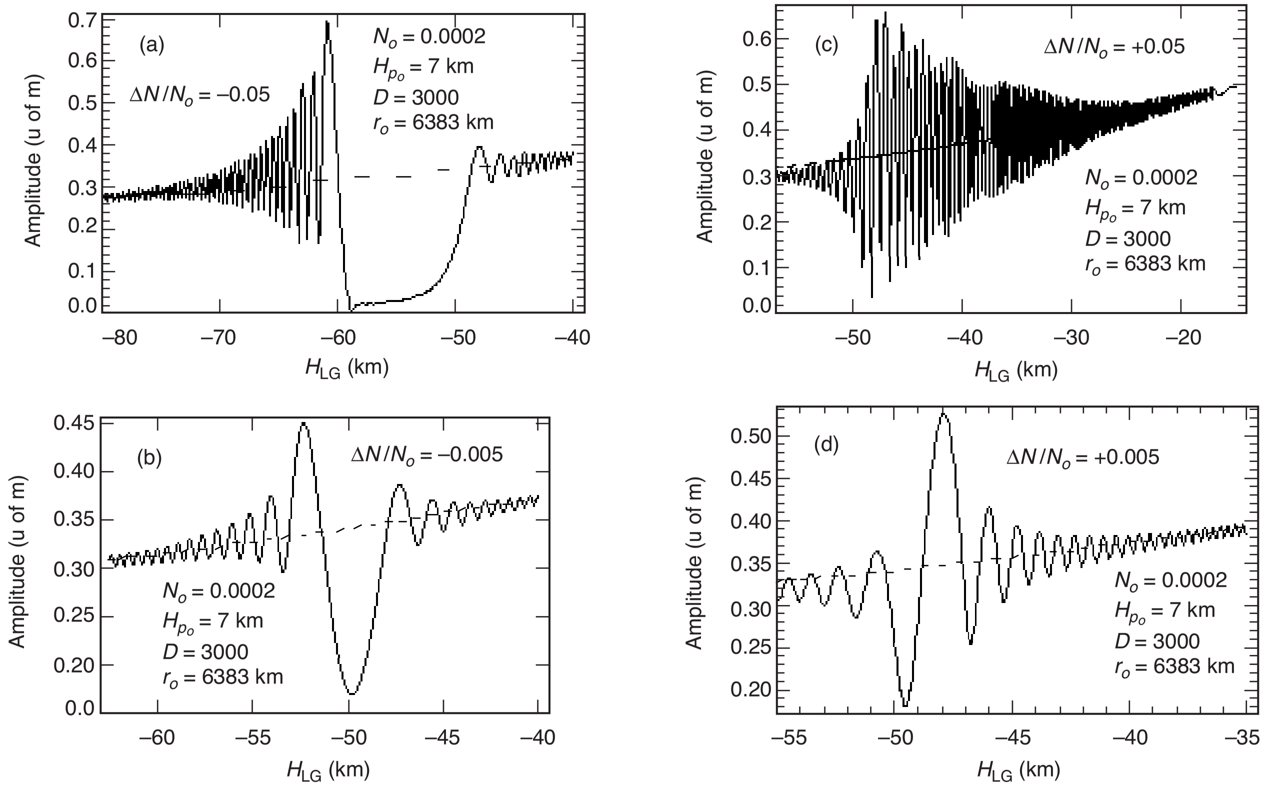

Fig. 2-11. Fresnel amplitude and phase perturbations at the LEO caused by a discontinuity in refractivity ($\Delta N = N_o^+ - N_o^-$) at about 4-km altitude: amplitude perturbations with (a) $\Delta N/N_o = -0.05$, (b) $\Delta N/N_o = -0.005$, (c) $\Delta N/N_o = +0.05$, and (d) $\Delta N/N_o = +0.005$, and phase perturbations minus geometric optics terms with (e) $\Delta N/N_o = -0.05$ and (f) $\Delta N/N_o = +0.05$. Super-refractivity occurs when $\Delta N < 0$. The dashed curves in (a) through (d) are from the defocusing factor for the dry-air exponential model used in the reference refractivity profile.

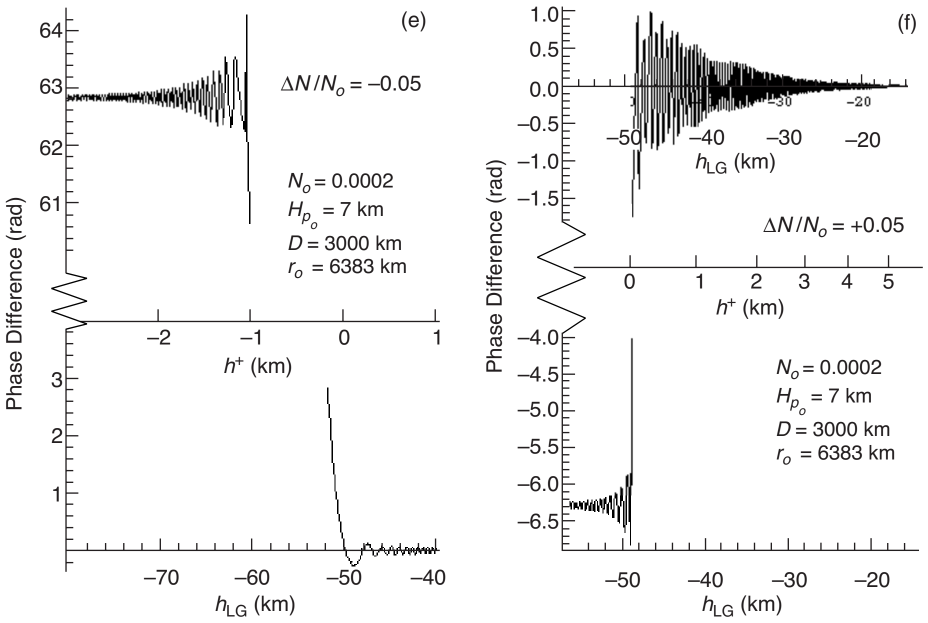

Fig. 2-11 (cont'd.).

2.7.1 Fresnel Response to a Discontinuity in Lapse Rate

Figures 2-10(a) through 2-10(d) show the Fresnel effects in phase and amplitude. The abscissa h_{LG} has the favorable property of varying nearly linearly with time during an occultation (within $\pm 0.5\%$ for the spans in these figures). The equivalent elapsed time in these figures is 5 to 7 s. The ray path altitude in Fig. 2-10 ranges from ~6 km to ~12 km and its scale, if shown, would be non-linear, and it would be compressed by defocusing compared to the h_{LG} scale. The defocusing here is about 1/3. The corresponding altitudes of the points of tangency in the actual atmosphere would be 400 to 500 m lower (essentially $(n-1)r_o$ lower). The boundary marking the discontinuity in lapse rate, at a thin-screen altitude h_o of 10 km above sea level, is crossed when the altitude of the LEO–GPS straight line reaches the altitude $h_{\mathrm{LG}} = h_{\mathrm{LG}}(1) = -5.78$ km. The dry air exponential refractivity model was used to generate the reference bending angle from Eq. (2.3-17) and the reference phase delay; it yields at $h_o = 10\,\mathrm{km}$ a reference bending angle of 5.44 mrad. The pressure scale height is 7 km, and the temperature is 210 K. At a LEO orbital altitude of 700 km, the limb distance is $D = 3000\,\mathrm{km}$. The principal Fresnel effects in phase and amplitude are completed within a time interval of 2 to 3 s.

The LEO-observed phase delay from the geometric optics phase delay (including the perturbed part below the boundary) has been suppressed in

Figs. 2-10(a) and 2-10(b) to stop the phase windup. This leaves only the Fresnel effects. For ray path altitudes below h_o, the question arises when a caustic is present (whenever $\Delta\gamma > 0$) as to which of the new rays should be used for phase-stopping. The geometric optics phase delay offset Δ^- from the phase delay of the main ray (m), or + ray, is shown versus h_{LG} in Fig. 2–12. This is the difference in phase of a– regime ray at the LEO relative to the phase of the + regime ray that either exists for $h_{\mathrm{LG}} \geq h_{\mathrm{LG}}(1)$ or would exist if the + regime were continued below the discontinuity. Figure 2–12 shows the phase offset Δ^- for the two rays in the multipath zone: the anomalous a ray and the branching b ray. The quantity Δ^{a} gives the phase offset for the a ray. This corresponds to the a ray in Fig. 2–2(a), which as a function of h_{LG} begins at the cusp at an

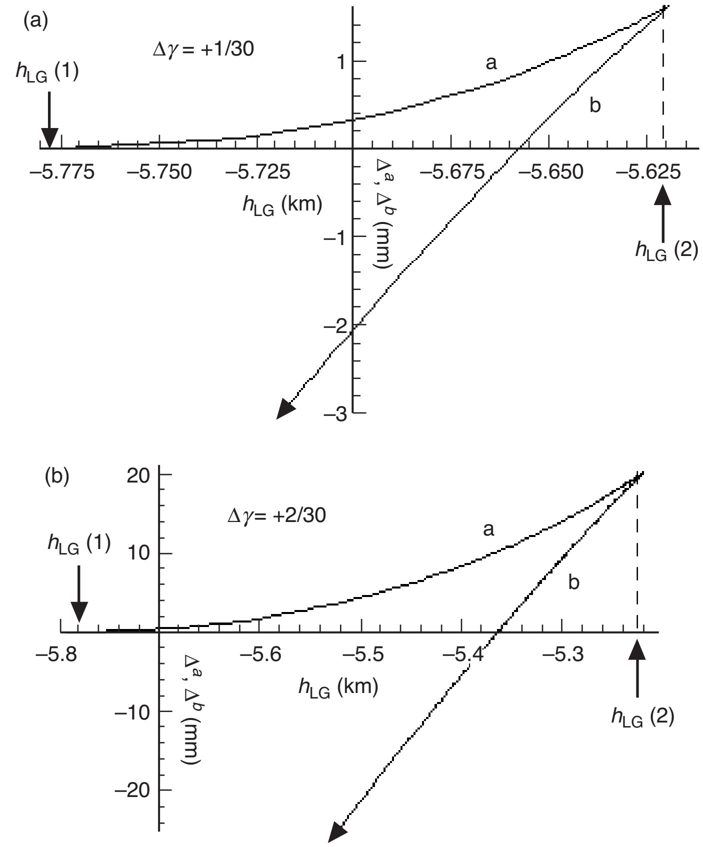

Fig. 2-12. LEO phase delay offsets from the main ray for Case C: (a) $\Delta\gamma$ = 1/30 and (b) $\Delta\gamma$ = 2/30. Phase offsets are for the ordinary branching ray, b, and also for the anomalous ray, a. Cusp marks the contact point with the caustic and its offset scales roughly as $(H_{p_o}\Delta\gamma)^4$. Delay is in range units for L1.

altitude of $h^a[h_{\mathrm{LG}}(2)]$ and terminates at the higher altitude $h_o = h^a[h_{\mathrm{LG}}(1)]$. This point $h_o = h^a[h_{\mathrm{LG}}(1)]$ also marks the termination of the lower descent of the + ray. Δ^{b} gives the phase offset for the b ray. It continues downward indefinitely into the – regime. The phase from the b ray has been applied in Fig. 2-10 to stop the phase windup. Consequently, there is a discontinuity in phase in Figs. 2-10(a) and 2-10(b) at $h_{\mathrm{LG}} = h_{\mathrm{LG}}(2)$. The magnitude of the discontinuity is about 1.6 mm for $\Delta\gamma = +1/30$ and about 20 mm for $\Delta\gamma = +2/30$.

Figures 2–10(a) and 2–10(c) show the Fresnel response for $\Delta\gamma = 1/30 > 0$. The flaring in amplitude in Fig. 2-10(a) at $h_{\mathrm{LG}} \cong -6\,\mathrm{km}$ and $h^+ \cong 9.8\,\mathrm{km}$ corresponds to the first contact with the caustic surface. In a descending occultation, the + ray at the altitude $h^+[h_{\mathrm{LG}}(2)]$, which is above h_o, abruptly splits into three rays, two of which, the a and b rays in the – regime, arrive at the LEO essentially mutually in phase initially from an altitude slightly below h_o ($h^-[h_{\mathrm{LG}}(2)] - h_o \approx -40\,\mathrm{m}$). The third, the + ray, which is the main ray or the m ray in Fig. 2–2, arrives from the original altitude above h_o, ($h^+[h_{\mathrm{LG}}(2)] - h_o \approx +50\,\mathrm{m}$) with an initial phase offset from the other two that is given by the cusp offset in Fig. 2–12(a).

The cusps in Fig. 2–12 mark the initial contact with the caustic surface for a descending occultation and also the point of nascence of the a and b rays. The altitude of the ray at the cusp is determined in terms of the Fresnel function $\Phi(h, h_{\mathrm{LG}})$ by the conditions that both $\partial\Phi/\partial h$ and $\partial^2\Phi/\partial h^2$ are zero there. The magnitude of the initial phase offset at the cusp is very sensitive to $\Delta\gamma$. It can be shown from Appendix C that this phase offset generated by a discontinuous lapse rate can be evaluated in terms of a power series in $\Delta\gamma H_{p_o}$ and that the leading term is $(\Delta\gamma H_{p_o})^4$. The relative magnitude of the flaring at the caustic contact point will depend on how flat the Fresnel phase function is in that neighborhood. This is discussed further in Appendix D.

If the phase offset at the cusp from the phase of the + ray approaches a radian, as is nearly the case for Fig. 2–12(b) (which corresponds to an unrealistically large lapse rate discontinuity of +14K/km), then a tracking-loop-driven GPS receiver would begin to encounter difficulty in such a neighborhood. The concomitant flaring at the caustic contact would likely lure the receiver to begin tracking one or the other of the nascent rays, a or b, there. These figures (Figs. 2-10(a), 2-10(b), and 2-12) also suggest that significant Fresnel ringing in phase and amplitude will persist well below the site of the discontinuity but that only faint Fresnel perturbations are evident for ray path altitudes higher than a Fresnel radius above the discontinuity when $\Delta\gamma > 0$. Asymptotic Fresnel effects typically attenuate slowly in their continuation past a discontinuity.

A lapse rate change of $\Delta\gamma = +2/30$ is too large for the tropopause and also one does not find a strict discontinuity there. Moreover, the tracking statistics through the tropopause from more recent LEOs is good. On the other hand, sharp changes in refractive gradient due to a water vapor layer in the lower troposphere or due to sharp electron density changes in the lower ionosphere (for example, across the boundary of a sporadic E-layer) can and do cause tracking difficulty for grazing ray paths. Multipath and diffraction effects for these situations can be quite severe, as they are extremely sensitive to the magnitude of the change in refractive gradient. The severity of the Fresnel phase perturbations in the vicinity of a discontinuity in scale height is, as in Case C, very sensitive to the magnitude of ΔH_{p_o}, roughly the phase perturbations scale as $(\Delta H_{p_o})^4$. Moreover, from Figs. 2–6(b) and 2-6(c) and from Appendix C, one concludes that the profile of $\delta\alpha(r)$ for Case B will be similar to that for Case C when $\Delta H_{p_o} / H_{p_o} \approx -\Delta\gamma H_{p_o}$. For example, it is shown from Appendix A that a value $\Delta H_{p_o} / H_{p_o} = -(\Delta\gamma H_{p_o})(H_{\rho_o} / H_{p_o})^2$ gives for Cases B and C the same value for the discontinuity in the refractive gradient at r_o; only the subsequent bending-angle profiles differ, but only slightly. When this correspondence between ΔH_{p_o} and $\Delta\gamma$ holds, one would expect to see Fresnel perturbations for Case B that are similar to those for Case C.

Figures 2–10(b) and 2-10(d) show examples of Fresnel effects when $\Delta\gamma < 0$. Here the caustic-induced flaring is absent. The Fresnel effects are smaller and their onset is lowered in altitude to h_o. Also, the severe darkening at $h = h_o$ predicted by Eq. (2.2–6) and shown in Fig. 2–8(a) is very prominent but modulated by diffraction.

2.7.2 Fresnel Response to a Discontinuity in Refractivity

Figure 2–11 shows the thin-screen/scalar diffraction prediction for a discontinuity in N itself at $r = r_o$, a Case A scenario. Here r_o corresponds to a lower troposphere altitude of about 1/2 scale height above the Earth's surface. The discontinuities are $\Delta N_o / N_o = \pm 0.05$ and $\Delta N_o / N_o = \pm 0.005$. The same exponential refractivity model given in Eq. (2.3–17) is used here for the reference bending angle and Fresnel phase function with $N_o = 0.0002$ and $H = 7\text{ km}$. The dashed curves in these figures are the defocusing from the bending-angle profile for the exponential reference refractivity based on geometric optics. The amplitude of the incident wave just before encountering the atmosphere is normalized to unity.

Figures 2-11(a) and 2-11(b) show the Fresnel response in amplitude to a negative discontinuous change $\Delta N = N^{+} - N^{-} < 0$. These can provide limiting forms for the amplitude perturbations experienced by a ray encountering from above a sharp water vapor layer at an altitude of about 4 km. A 5 percent discontinuity in refractivity for these physical conditions corresponds roughly to only a 10 to 20 percent change in water vapor density. In Appendix C, the sensitivity of the corresponding Doppler spread between multiple rays to the magnitude of $\Delta N_o / N_o$ for $\Delta N_o < 0$ is discussed (see Fig. C–1).

Figures 2-11(c) and 2-11(d), which show the Fresnel response to a positive discontinuity in N, provide examples of what might be observed as the tangency point of the ray drops below the bottom of such a water vapor layer and enters a regime of drier and speedier air. This corresponds to the ray diagram in Fig. 2–2(a).

A relatively small negative change in refractivity, 5 percent for Fig. 2-11(a), leads to a very deep shadow zone lasting roughly 4 s. In Appendix C, it is shown that the shadow zone duration is closely proportional to $(-\Delta N)^{1/3}$. Thus, even very small negative discontinuities can result in significant shadow zones. In Fig. 2-11(b), the magnitude of ΔN is a factor of ten smaller than that used in Fig. 2-11 (a).

On the other hand, a relatively small positive ΔN results in significant interference well above the boundary. It is shown in Appendix C that the duration of the multipath region for a positive discontinuity is closely proportional to $(\Delta N)^{1/2}$. Therefore, the maximum altitude separation of the rays and also the Doppler spread for a positive discontinuity is approximately proportional to $(\Delta N)^{1/2}$.

The super-refractivity condition from a discontinuity in N creates errors in our thin-screen/scalar diffraction prediction for observed phase. The phase, unlike the amplitude, is extremely sensitive to model approximations. The uniqueness condition, $1 + a \tan\alpha (d\alpha / da) > 0$, is effectively violated across the discontinuity in the bending angle (the Snell term). Chapter 3, which uses Mie scattering theory to predict phase and amplitude at the LEO, also compares those wave theory results with the thin-screen results for similar values of ΔN (see Figs. 3–23 through 3–26).

One example of the phase perturbations from a discontinuity in N is shown in Figs. 2–11(e) and 2–11(f) in order to point out that the phase effects across a discontinuity depend strongly on the polarity of ΔN. This also is borne out in the Mie scattering treatment in Chapter 3. In Figs. 2–11(e) and 2–11(f), the phase windup has been stopped by subtracting from the predicted phase observed at the LEO the geometric optics phase applicable to each regime, the "+" and "–" regimes, as already explained earlier for Fig. 2–10. In Fig. 2-11(e), where $\Delta N / N = -0.05$, we have subtracted the geometric optics phase profile using the reference bending angle model at all altitudes, including across the shadow zone. However, no rays exist in the shadow zone [see Fig. 2–7(a)] because of the discontinuity in refractivity. For Fig. 2-11(e), the shadow zone is defined by $h_{\mathrm{LG}}(2) \le h_{\mathrm{LG}} \le h_{\mathrm{LG}}(1)$; $h_{\mathrm{LG}}(2) \approx -60.5\ \mathrm{km}$, and $h_{\mathrm{LG}}(1) \approx -49.3\ \mathrm{km}$, a temporal separation of 3 to 4 s. The shadow zone marks a transition region where the Fresnel phase migrates from the reference Fresnel phase profile in the + regime to the perturbed Fresnel phase profile in the − regime. The latter involves (in a geometric optics context) the weighted (by their respective defocusing factors) vector addition of the phase contributions from both the b and a rays. The point $h_{\mathrm{LG}}(2) \approx -60.5\ \mathrm{km}$ marks first contact with the caustic. Below this point, we have subtracted the phase from the branching ray b. Therefore, across this transition region we see cycle slipping relative to the prediction provided by the reference Fresnel phase profile, which here is the extrapolated reference profile from the main ray that would apply in the shadow zone without the discontinuity. The cycle slip essentially equals the difference between the phase of the nascent rays at the caustic contact point at the end of the shadow zone at $h_{\mathrm{LG}} = h_{\mathrm{LG}}(2)$, minus the phase of the main ray extended to that caustic point, even though the main ray doesn't exist there. For the example in Fig. 2-11(e), this amounts to about 10 cycles. For the opposite polarity of the same magnitude, i.e., $\Delta N / N = +0.05$ shown in Fig. 2-11(f), which involves a positive ΔN of the same magnitude, the cycle slip is an order of magnitude smaller.

The ramification of this for a statistical net loss of cycles by a receiver attempting to track a GPS signal that passes through a series of sharp positive and negative transitions in refractivity should be noted. A GPS receiver with an ordinary low-gain antenna is likely to have more difficulty tracking through a deep trough in the shadow zone because of poorer signal-to-noise ratio (SNR), as shown in Fig. 2–11(a), than it would in a strong (but messy) signal condition, as shown in Fig. 2–11(c). Note from Figs. 2–11(e) and Fig. 2–11(f) the difference in the rate of unaccounted-for phase windup between these positive and negative ΔN scenarios. Statistically this could result in a net loss of cycles incurred by the receiver in attempting to track the phase during successive episodes of sharp transitions of positive and negative ΔN. A net loss of cycles leads to an underestimate of N.

The frequency spectrum of the fringes in Fig. 2–11(e) is biased significantly higher than that for Fig. 2–11(f), even though the magnitude of the discontinuity in refractivity is the same in the two sets of figures. This is because of the on-average wider separation of the multipath tones when $\Delta N > 0$ (see Appendix C, Fig. C–2). The multipath separation altitudes in the panels for $\Delta N > 0$ are sufficiently wide compared to the vertical width of the first Fresnel zone that a geometric optics approach to construct these figures using Eqs. (2.2–5), (2.2–6), and (2.4–1) should be fairly accurate. Also, no caustics occur in the neighborhood about these altitudes in this case. Use of geometric optics is, of course, much easier when corrections for diffraction can be neglected. On the other hand, for the $\Delta N < 0$ case shown in Figs. 2–11(a) and 2–11(b), the dire predictions of geometric optics of infinite signal power at the contact point with the caustic surface and zero power in the shadow zone are far from accurate. Ray theory based upon second-order theory fails in these regions, or at least it is compromised. In the former region, it fails because $dh_{\mathrm{LG}} / dh \to 0$ at the caustic contact point and, therefore, $\partial^2 \Phi(h, h_{\mathrm{LG}}) / \partial h^2 = 0$ at that point, rendering the quadratic form of the stationary-phase technique useless. In the latter region, it fails because $\partial \Phi(h, h_{\mathrm{LG}}) / \partial h \neq 0$ at any altitude; there are no stationary-phase points there. The scalar diffraction integral does a good job with these two situations.

2.7.3 A Boundary Layer

An example of a two-sided boundary is shown in Figs. 2–13(a) through 2-13(d). This example might be found in a marine layer in the lower troposphere with very sharp boundaries. Here the refractivity, shown in Fig. 2–13(a), abruptly within 50 m rises by 5 percent over a very short range compared to the first Fresnel zone from the dry air reference refractivity profile used in Fig. 2–11. For this transition, $r_o n_o' + n_o = -0.2$; therefore, n_o' exceeds the super-refractivity limit of $-n_o / r_o \approx -1/6400\ \mathrm{km}^{-1}$. Then at an altitude of 1/2 kilometer lower, the refractivity abruptly drops by 5 percent. The bending-angle profile for this example is shown in Fig. 2–13(b). It is obtained from Eq. (2.2–2), taking care to avoid integrating across a very narrow super-refractivity interval near the upper boundary where $nr_* < a$. Figure 2-13(c) shows the (h, h_{LG}) relationship using the thin-screen model. In this figure, the curve for h_{LG} versus h from the reference refractivity profile is dashed below the upper boundary. Figure 2-13(d) shows the Fresnel response in amplitude for this model. The dashed curve in this figure is the defocusing from the dry air reference refractivity profile. Qualitatively, this figure is largely a composite of Figs. 2–11(a) and 2–11(c). The shadow zone due to the strong refractivity gradient over the transition at the upper boundary (where $dN / dh < 0$) is largely filled in by the "throw-back" rays from the reverse transition ($dN / dh > 0$) at the lower boundary. It goes almost without saying that the refractivity transitions across boundary layers in the real atmosphere can and do produce very complicated interference/diffraction patterns in amplitude and phase, such as those shown in Figs. 1–3 and 1–8.

Higher SNR will be called for in future missions if the fine structure in the narrow refracting layers is to be fully studied to their sub-Fresnel resolution limit. This means, for example, the use of an antenna array that is effectively several decimeters in size along the vertical dimension, in-phase and quadrature dual-band coarse acquisition (C/A) recovery of carrier phase, and so on. Sample rates somewhat higher than the canonical 50-Hz GPS data transmission chip rate may prove beneficial.

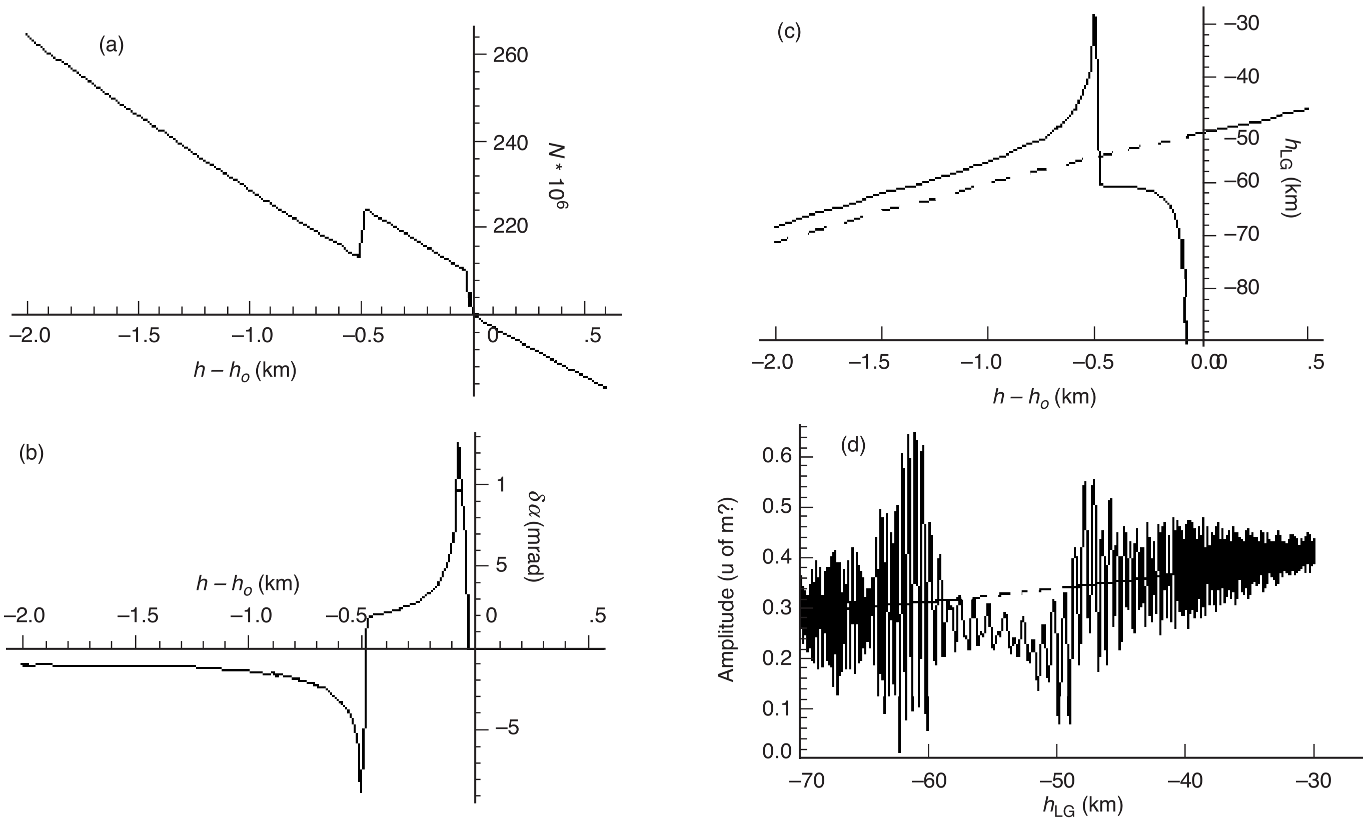

Fig. 2-13. Boundary layer at 4-km altitude with a sharp transition in refractivity of 5 percent at the upper and lower boundaries: (a) the refractivity profile, (b) the resulting bending angle perturbation, (c) the thin-screen relationship, and (d) the amplitude perturbations. The same reference refractivity model was used in Fig. 2-11. The changes in N occur over a transition interval of 50 m at each boundary. The upper boundary is super-refractive, $r_o n_o' = -1.3$.

2.8 Sensing a Boundary in the Ionosphere

In the ionosphere, the plasma frequency is typically about 3 MHz. Hence, the ionosphere, except on occasions of severe storms resulting in deep scintillation, is essentially transparent for radio signals at L–band frequencies, with typical lower ionosphere refractivity in the $10^{-5} - 10^{-6}$ range. On the other hand, the Doppler and SNR signatures in observed GPS signals that traverse the ionosphere at grazing angles can reflect a high degree of spatial and temporal variability in electron density. An example is provided in Fig. 2–14 [19], which shows the vertical profiles of the L1 and L2 bending angles obtained from GPS/MET for a particular occultation on April 24, 1995. As the tangency points of the two rays descend through the lower ionosphere (~50 to 250 km) during this occultation, sharp changes in the vertical gradient of the electron density in this neighborhood can be inferred. Around 100 km these profiles reveal a layer with near-discontinuities in electron density across their upper and lower boundaries. Figure 1–3 also shows a sharp transient in $\mathrm{SNR_V}$ for this same occultation around 100-km altitude.

The GPS/MET observations for certain occultation profiles show a near blackout or fadeout in SNR in the GPS signal when the point of tangency of the ray at the Earth's limb drops through this kind of boundary in the ionosphere. Figure 2–15 [23] captures in fine detail an example of this fading (associated with a sporadic E-layer) in the same occultation taken on April 25, 1995. It charts at a 50–Hz sample rate the variation of the voltage signal-to-noise ratio

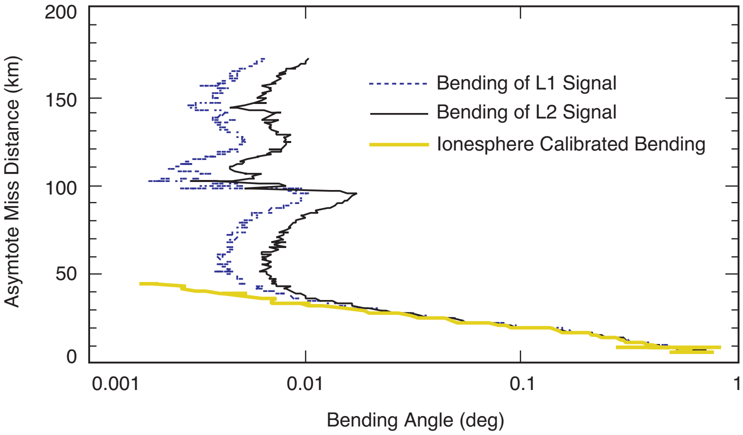

Fig. 2-14. Example of GPS/MET-observed bending angle profiles for the L1 and L2 signals, and also the "ionosphere-free" profile. Redrawn from [18].

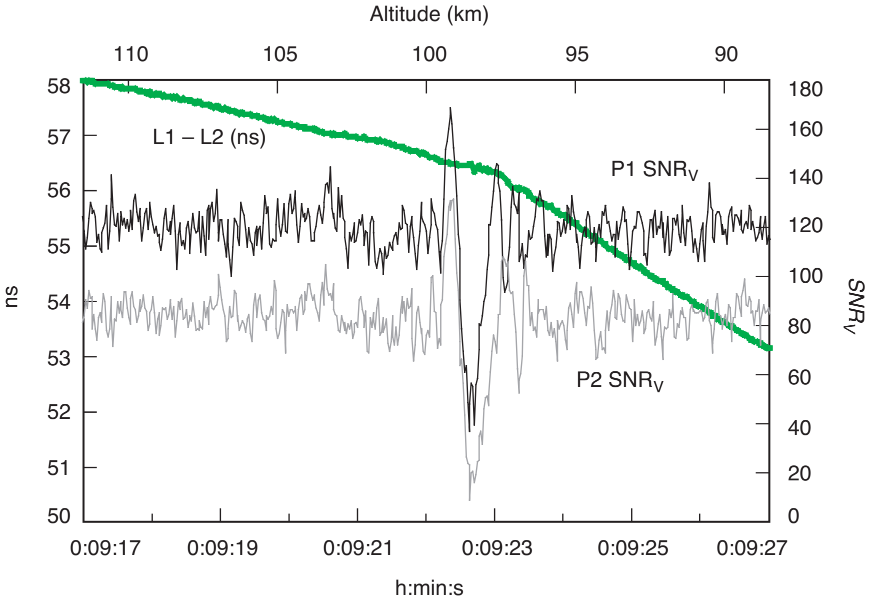

Fig. 2-15. Signal amplitude profiles over a 10-s interval for the L1 and L2 carriers observed by GPS/MET for an occultation on April 25, 1995, sampled at 50 Hz. Amplitude is voltage signal-to-noise ratio SNR$_V$ **at 1 s. The phase difference of the L1 and L2 carriers is in nanoseconds. The 100-km altitude of the transient indicates the crossing of a sporadic E-layer. Redrawn from [23].**

SNR$_V$, rated 1-s average,[4] for the L1 and L2 carriers over a 10-s time interval; the variability in the two SNR$_V$ profiles is highly correlated, reflecting true variability in electron density. This figure also shows the change in the difference of the observed L1 and L2 carrier phase [L1–L2] with time (each carrier phase in nanosecond units).

In Fig. 2-16, the excess Doppler for the L1 and L2 carriers for this occultation is shown at 50-Hz and 5-Hz sample rates [24]. The latter is obtained by applying a sliding box filter of 200-ms width to the 50-Hz series. A value of 100 for SNR$_V$ corresponds to a thermal noise in an individual L1 phase measurement over a 20-ms average of about 2 mm. The thermal noise on most 50-Hz Doppler points in Fig. 2-16 is about 0.7 Hz on L1 and 1.0 Hz on L2. Below $\mathrm{SNR_V} \cong 30$ the receiver is increasingly likely to encounter difficulty tracking the phase, particularly when significant phase accelerations are also present. On this particular occasion, the receiver maintained lock.

---

[4] Here SNR$_v$ is expressed as the ratio of the signal amplitude to noise amplitude in voltage that would be obtained by averaging over 1 s, even though the actual averaging time in Figs. 2-15 and 2-16 is 20 ms, the reciprocal of the sample rate.

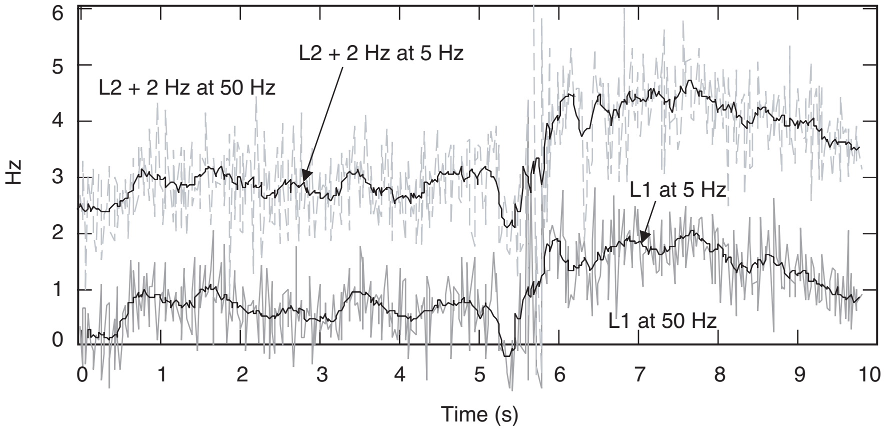

Fig. 2-16. The excess Doppler profile for the same time interval shown in Fig. 2-15. Thermal noise sampled at 50 Hz is about 0.7 Hz (1 σ) on the L1 Doppler and 1.0 Hz for L2. The 5-Hz series are obtained by averaging the 50-Hz points over 200 ms. Redrawn from [23,24].

The Doppler change is greater for L2 after crossing the "boundary" at approximately $t = 5.7$ s. The dip in Doppler just prior to impacting the boundary indicates a negative radial gradient in the electron density transient above the boundary. The abrupt rise in Doppler just afterwards indicates a large positive gradient below.

The precipitous fade and recovery of $\mathrm{SNR_V}$ in Fig. 2-15 and the abrupt change in slope of [L1–L2], all occurring within about 1/3 s about coordinated Universal time (UTC) = 00:09:22.7, mark a lower boundary of an ionospheric layer of higher electron density. For this occultation, a very abrupt roll-off of electron density occurs with descending altitude. The scenario shown in Fig. 2-2(c) best matches this situation.

The measured $\mathrm{SNR_V}$ and Doppler profiles on L1 and L2 each provide a number of key quantities: the change in Doppler across the fadeout zone (and hence the change in bending angle), the epoch of onset (and hence altitude from POD information and ray tracing) of the fadeout (which may differ on L1 from that on L2 because of ray splitting), the widths of the fadeout zones, and their depths. For the example in Fig. 2-15, the difference in onset of the L1 and L2 rays is about 0.1 s or ~200 m. Using these observed quantities and a model, one can recover, in addition to the altitude of the transition boundary, the change in electron density at the boundary, its gradients on each side of the boundary (with an additional global ionosphere model), the mean distance D of the layer from the LEO, and the separation of the L1 and L2 rays at the boundary.

2.8.1 Fresnel Effects at a Boundary

An example of this approach and a discussion of some Fresnel features are now given. Let the boundary be located at $r = r_o$. We present here the response of the ray in terms of the LEO-observed phase and amplitude using a simple *locally* spherical symmetric model in which the electron density undergoes a sharp change at the boundary but is otherwise smooth. For calculating the change in bending angle due to this roll-off in electron density, the assumption of local spherical symmetry should be useful because of the abrupt nature of the roll-off and the narrowness of its extent in the radial direction. (For the ionosphere at large, the assumption of spherical symmetry has been shown to be unreliable [19].) We will obtain an estimate of the change in electron density profile by assuming that over a short altitude range around $r = r_o$ (roughly 1 to 2 km) the variations in bending angle due to the ionosphere at large can be ignored. We also assume here that the radius of curvature of the ray is very much larger than the local radius of curvature of the boundary, and that the latter equals r_o and is radial directed.

The refractivity ($\times 10^{-6}$) for carrier phase in the ionosphere is given by the principal term

$$N = -\kappa \frac{n_e}{f^2} \tag{2.8-1}$$

where n_e is the electron density, f is the carrier frequency, and $\kappa = 40.3\,\mathrm{m}^3 \cdot s^{-2} \cdot e^{-1}$. A fuller description of ionospheric refractivity, including the higher-order terms, is given in [4,20,21]. Let Δn_e be the change in the electron density in the neighborhood of r_o relative to a reference electron density profile for the ionosphere at large that is assumed to change more gradually with radial distance. For one example of a global reference model, see Appendix E. We assume that the spatial variability of the reference profile is small compared to the magnitude of the local electron density gradient at the roll-off. Inasmuch as the entire fadeout zone in Fig. 2–15 spans only about 1/3 s of time, the boundary width is comparable to or less than the first Fresnel zone for this example; a Fresnel treatment will be needed to predict the observed phase and amplitude. Under these conditions, the bending-angle profile resulting from a discontinuous change in refractivity, as provided by Case A, should be a useful initial approach. The change $\Delta\alpha_S$ below the boundary $r = r_o$ for this case is given by the discrete Snell term from Eq. (2.3–27) by

$$\Delta\alpha_S(r) = \left\{\begin{array}{l} 2\left(\sin^{-1}\left(\dfrac{n^- r}{n^+ r_o}\right) - \sin^{-1}\left(\dfrac{r}{r_o}\right)\right) \doteq 2\sqrt{2}\left(\sqrt{v} - \sqrt{\Delta N + v}\right),\ r \le \tilde{r} < r_o \\ 0,\ r > r_o \end{array}\right\} \quad (2.8\text{-}2)$$

where[5] $v = (r_o - r)/r_o = (h_o - h)/r_o$ and $n^+ = n^- + \Delta N$. Here ΔN is obtained from Eq. (2.8–1) and it is a negative quantity when $\Delta n_e > 0$; i.e., when the electron density increases upward across the boundary. The quantity $\tilde{r}$ marks the upper limit of the ray due to the internal critical refraction limit when $\Delta N < 0$; it is defined by the condition $n^-\tilde{r} = n^+ r_o$. When $r \le \tilde{r}$, the signal while traveling through the lower region experiences a slowdown in phase velocity due to the increased number of electron "holes" with depth in that region. For $r > r_o$, $\Delta\alpha(r) = 0$.

Using the thin-screen model and the stationary-phase condition, it follows that the phase delay $\varepsilon(h) = \delta\varphi(h)$ embedded in the thin screen is given by

$$\varepsilon(h) = \left\{\begin{array}{l} k\displaystyle\int_r^{\tilde{r}} \Delta\alpha_S(r')dr' \doteq \frac{2^{5/2}}{3} k r_o\left(v^{3/2} - (-\Delta N)^{3/2} - (\Delta N + v)^{3/2}\right), r < \tilde{r} \\ 0,\ r > \tilde{r},\quad v = (r_o - r)/r_o = (h_o - h)/r_o \end{array}\right\} \quad (2.8\text{-}3)$$

Figures 2–17 through 2-19 show the change in observed phase and amplitude that result from the diffraction integral in Eq. (2.5–1) using this Case A model. For the L2 carrier, a discontinuous electron density of $\Delta n_e = +4.48 x 10^{10}$ e/m^3 corresponds to $\Delta N_{\mathrm{L}2} = -1.2 \times 10^{-6}$. The abscissa $r_o - r_{\mathrm{LG}}$ is the difference in altitudes of the boundary and the tangent point of the straight line between GPS/MET and the occulted GPS satellite; $r_o - r_{\mathrm{LG}}$ varies nearly linearly with time. Depending on the averaging time used to construct the phase measurements (and hence Doppler), Fig. 2–17 shows that the Fresnel perturbations[6] can be a significant error source for sharp transitions in electron density if unaccounted for in the inversion process that converts Doppler into an electron density.

---

[5] For notational convenience in this section we have dropped the subscript "$_*$" from r_*, the turning point of the ray, i.e., $r \equiv r_*$. Also, in the ionosphere, we can set $r - r_o = h - h_o$ because the nominal defocusing is minimal.

[6] The Fresnel perturbation is the difference between the phase profile shown in Fig. 2–17 minus the profile from geometric optics, which is given by $\Delta = 1/2kD\left(\Delta\alpha_S(v)\right)^2 + \varepsilon(v),\ r_o v = r_b + D\Delta\alpha_S(v)$.

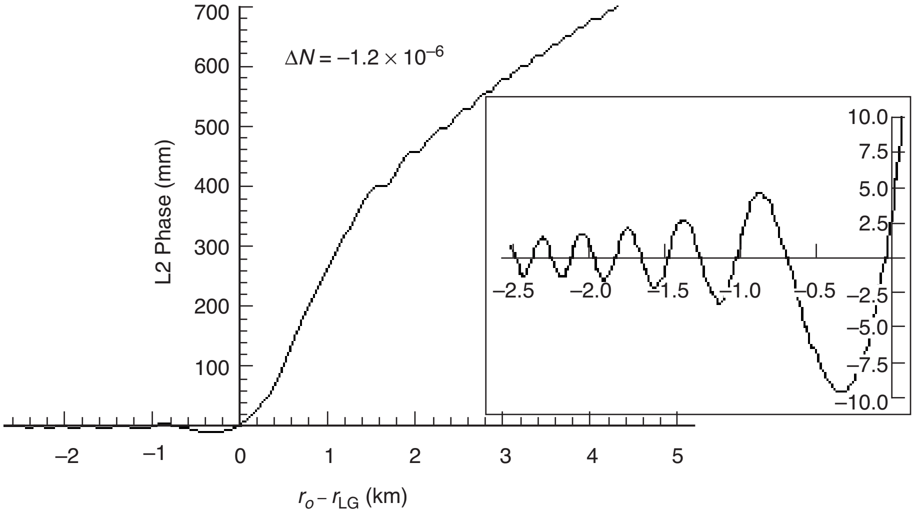

Fig. 2-17. Fresnel perturbations in phase at the LEO from a discontinuity $\Delta N_{L2} = N_{L2}^{+} - N_{L2}^{-} = -1.2 \times 10^{-6}$ **at 100-km altitude. For L2, this is a change in electron density of** $\Delta n_e = +4.48 \times 10^{10}$ e/m^3**. Inset shows perturbations for** h_{LG} **just above the boundary.**

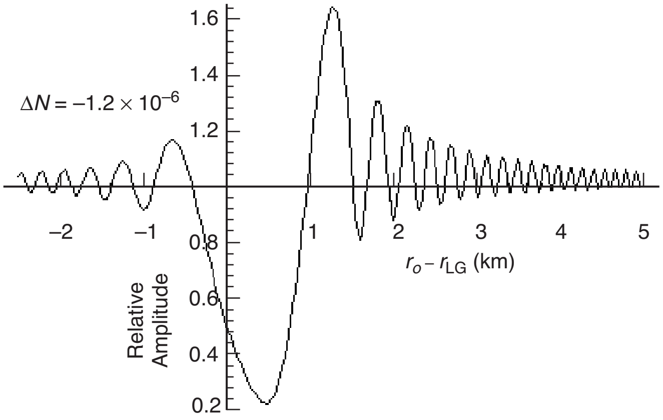

Fig. 2-18. Fresnel perturbation in L2 signal amplitude due to an upward discontinuity $\Delta n_e = +4.48 \times 10^{10}$ e/m^3 **at 100-km altitude.**

Figure 2-19 shows the Fresnel difference $L1 - L2$, each carrier measured in range units (ns) using the Case A model with $\Delta n_e = +4.48 x 10^{10}$ e/m^3. Figures 2-17 through 2-19 span about 3 s of time. For the specific occultation shown in Figs. 2-15 and 2-16, $\dot{r}_{LG} = -2.3$ km/s.

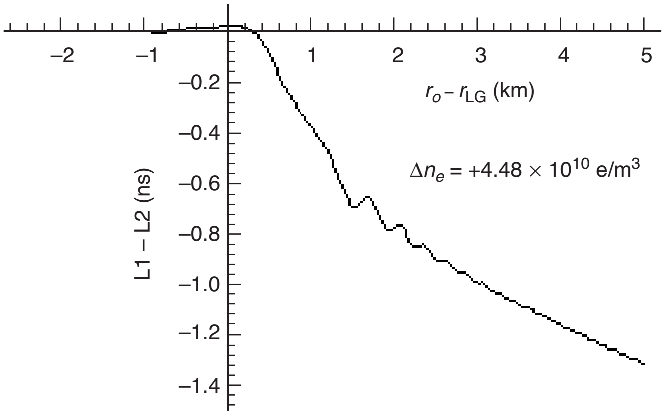

Fig. 2-19. Differenced L1–L2 phase perturbation due to a discontinuity Δn_e, at 100-km altitude.

Above the boundary, the L1–L2 profile for the GPS/MET occultation shown in Fig. 2-15 is sloped downward at an average rate of about –0.28 ns/s (which reflects the presence of a more global electron density distribution above the boundary, and which yields average L1 and L2 bending angles of about +0.05 and +0.08 mrad, respectively). Below the boundary, the average slope is about –0.85 ns/s. Thus, the change in average slope as the ray cuts down across the boundary and into the lower region is about –0.57 ns/s. For the small angles involved here, the scale of the observed change in phase delay is nearly proportional to the change in average slope of L1–L2 in Fig. 2-15. One can obtain an estimate of Δn_e by adjusting its value used to generate Fig. 2-19 to obtain a match in the average slope differences of the L1–L2 profiles across the boundary in Figs. 2-15 and 2-19.

2.8.2 Amplitude Effects at a Boundary

The widths of the fadeout zones in $\mathrm{SNR_V}$, shown in Fig. 2-15 for L1 and L2, provide additional information. Let B be a measure of the width in r_{LG}-space across the SNR fadeout zone. A simple measure is the distance between the boundary and the altitude of the first contact with the caustic, near which the flaring would be at maximum. Figure 2-18 shows the Fresnel perturbation in observed signal amplitude using the Case A model with $\Delta N_{\mathrm{L}2} = -1.2\times 10^{-6}$. However, we note from this figure that the Fresnel perturbations on the SNR tend to soften the response of the observed signal to the discontinuity. This somewhat fills in the trough in the SNR profile and widens the distance between the point of maximum Fresnel overshoot just above the boundary and the point of maximum flaring, due to the caustic, just below the boundary. See Appendix D for a discussion of Fresnel effects on location and magnitude of the flaring. For the value of Δn_e used in Fig. 2–18, the altitude in r_{LG}-space of the first contact point with the caustic based on geometric optics is +1.05 km (see Appendix C). The actual maximum in Fig. 2–18 occurs at about +1.3 km, while the maximum Fresnel overshoot occurs at –0.7 km. Thus, this particular measure for B (from boundary to contact) will underestimate the width, more like a half-width. A correction to B for Fresnel effects can be developed. Asymptotic expressions for the Fresnel integrals can be used to isolate the point of maximum overshoot, and Appendix D provides Fresnel corrections to the position of maximum flaring and its maximum value.

From the thin-screen model using Eqs. (2.2–5) and (2.5–6), and using a power series expansion of Eq. (2.8–2) and the condition that $\partial h / \partial h_{\mathrm{LG}} \to \infty$ at first caustic contact, one obtains an expression for the "ray" offset $v^{\dagger} = \left(r_o - r^{\dagger}\right) / r_o = \Delta r_o / r_o$:

$$v^{\dagger} \doteq \frac{1}{2}\left(-\frac{2D\Delta N}{r_o}\right)^{2/3}, \quad v^{\dagger} >> |\Delta N| \tag{2.8-4}$$

For B, using the thin-screen model, one obtains

$$B = r_o - r_{\mathrm{LG}}^{\dagger} = r_o v^{\dagger} + D\Delta\alpha\left(r^{\dagger}\right) \doteq \frac{3}{2} r_o \left(-2D\Delta N / r_o\right)^{2/3} = 3 r_o v^{\dagger} \tag{2.8-5}$$

Let $B_{\mathrm{L}1}$ and $B_{\mathrm{L}2}$ denote the widths for L1 and L2, respectively. Then it follows from Eqs. (2.8-1) and (2.8–5) that

$$\frac{B_{\mathrm{L}1}}{B_{\mathrm{L}2}} \doteq \left(\frac{f_{\mathrm{L}2}}{f_{\mathrm{L}1}}\right)^{4/3} \tag{2.8-6}$$

which appears to hold roughly for Fig. 2-15. From Eqs. (2.8-4) and (2.8-5), one obtains for Δr and B the values $\Delta r = 350\,\mathrm{m}$ and $B_{\mathrm{L}1} \approx 1.0\,\mathrm{km}$ for $\Delta N_{\mathrm{L}1} = -1.2 \times 10^{-6}$. For $\dot{h}_{\mathrm{LG}} = -2.3\,\mathrm{km/s}$, this yields a temporal width of about 1/3 s, which is about the observed value in Fig. 2-15 for the half-width.

Figure 2-18 also shows a small asymptotic bias in amplitude above unity for the lower altitude range, which is a lense effect caused by the positive bending angles below the boundary.

2.8.3 Ray Splitting

The information about the change in slopes of the L1 and L2 carrier phase profiles and about the fadeout width provides redundant determinations of Δn_e; one can use this redundancy to recover an estimate for D. Let p denote a parameter such as the RF carrier frequency; in this case $p = f^{-2}$. Let $\psi(r_p, r_{\mathrm{LG}})$ be the excess carrier phase observed by the LEO—that is, the difference of the observed phase minus the predicted phase in the absence of the atmosphere and ionosphere. Here $r_p - R_E$ is the altitude of the point of tangency with the Earth's limb for the ray path corresponding to the parameter value *p*; similarly, r_{LG} is the position of the straight line between the LEO and the observed GPS satellite at its point of tangency. We use the subscript *p* on the ray path tangency point to denote the dependence of r_p at a fixed epoch (and, therefore, for a given value of r_{LG} on the parameter *p*, that is, $r_p = r_p[p, r_{LG}]$). If one varies *p* while holding the end points fixed and hence r_{LG} fixed, then r_p also must vary in such a way that the new ray path satisfies the stationary-phase condition. The observable α_{p_k} is given to first order in α_{p_k} (Appendix A) from the excess Doppler by

$$\alpha_{p_k} = \alpha\left[p_k, r_{p_k}\right] \doteq -\frac{\lambda_k}{2\pi \dot{r}_{\mathrm{LG}}} \dot{\psi}\left(r_{p_k}, r_{\mathrm{LG}}\right) = -\frac{c}{\dot{r}_{\mathrm{LG}}} \dot{L}k \tag{2.8-7}$$

where $\mathrm{L}k$ is the *k*th carrier phase in time units and *c* is the velocity of light. We have written $\alpha_p = \alpha\left[p, r_p\right]$ to reflect the direct and indirect dependency of the bending angle on the parameter *p*. The quantities $\dot{\psi}$ and $\dot{r}_{\mathrm{LG}}$ are known from the observations, and the POD information is known for the occulted GPS satellite and for GPS/MET. Therefore, bending-angle profiles, such as those shown in Fig. 2-14, follow directly from the Doppler measurements and the POD information. The quantities $\dot{\psi}_{\mathrm{L}1}$ and $\dot{\psi}_{\mathrm{L}2}$ in hertz are shown in Fig. 2-16.

Because of ionosphere-induced ray splitting, the tangent points for the L1 and L2 rays will not quite be the same for a given epoch. For example, there is ~0.1 s difference in the epochs of onset of the L1 and L2 fadeout zones in Fig. 2-15. This reflects an altitude difference due to ray splitting from the ionosphere at large that is of the order of 100 m. Figure 2-20 shows the profile of the ray splitting term $D(\alpha_{\mathrm{L}1} - \alpha_{\mathrm{L}2})$ using Eq. (2.8-7) for the same occultation shown in Figs. 2-15 and 2-16. Figure 2-20 shows the profile for a 50-Hz sample rate, and it shows the profile for a 5-Hz sample rate, which is derived by applying a sliding box filter of 200-ms width to the 50-Hz time series. The magnitude of the thermal noise reduces as the square root of the ratio of the sample rates, but the magnitude of the Fresnel fringes, which are essentially sinusoidal over sampling widths used in Fig. 2-20, reduces linearly. After contacting this particular ionospheric layer, the ray splitting essentially triples in magnitude. We should account for this (usually) small difference in altitude or, equivalently, a bending-angle difference for this example of the order of a tenth of a milliradian.

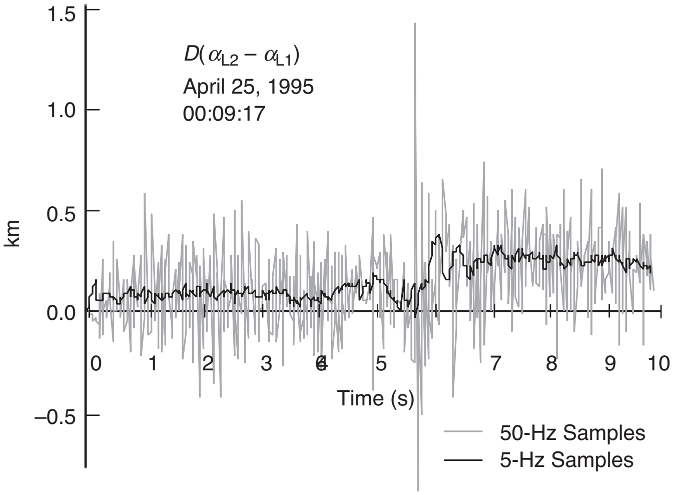

Fig. 2-20. Ray path splitting of the L1 and L2 signals for the same occultation and epoch shown in Figs. 2-15 and 2-16.

Reference [22] uses a linear combination of bending angles to eliminate ionosphere effects to second degree in frequency from the observations. We use a similar approach here to account for ray splitting. To a good approximation, the total bending can be decomposed into

$$\alpha_p = \alpha\left[p, r_p\right] = \alpha^A\left(r_p\right) + \alpha^I\left(p, r_p\right) \tag{2.8-8}$$

where α^A and α^I are, respectively, the atmospheric and ionospheric components of the bending. However, when the tangent point of the ray lies in the ionosphere, $\alpha^A \equiv 0$. Note that α^I depends on p directly [through the refractivity given by Eq. (2.8-1)] and also indirectly through r_p, which for a fixed epoch also depends on p. That is,

$$\alpha^I\left(p, r_p\right) \doteq p I\left(r_p\right) \tag{2.8-9}$$

where $I\left(r_p\right)$ is the path integral $\kappa \int \nabla n_e \times d\mathbf{s}$ along the ray through the ionosphere from the LEO to the GPS satellite with an impact parameter value r_p. For example, when the ray path tangency point lies in the ionosphere and if spherical symmetry were to apply,

$$I\left(r_p\right) \doteq \kappa r_p \left(\int_{r_p}^{r_{\mathrm{LEO}}} \frac{n_e'}{\sqrt{r^2 - r_p^2}} dr + \int_{r_p}^{\infty} \frac{n_e'}{\sqrt{r^2 - r_p^2}} dr \right) \tag{2.8-10}$$

Equation (2.8-10) has been simplified somewhat relative to Eq. (2.2-2) by setting the index of refraction terms in the integral to unity; in the upper stratosphere and the ionosphere, the differences in these terms from unity is very small. Also, the third- and fourth-order ionosphere terms have been ignored (except indirectly through ray splitting).

At a fixed epoch where r_{LG} is fixed, that is, when the end points of the ray are fixed, varying p will cause r_p also to vary so as to maintain the stationary-phase condition on the resulting ray path. Therefore,

$$\frac{d\alpha}{dp} = \frac{\partial \alpha}{\partial p} + \frac{\partial \alpha}{\partial r_p}\frac{dr_p}{dp}, \quad r_{\mathrm{LG}} \text{ fixed} \tag{2.8-11}$$

But using the thin-screen approximation [Eq. (2.2-5)],

$$\frac{dr_p}{dp} = D\left(\frac{\partial \alpha}{\partial p} + \frac{\partial \alpha}{\partial r_p}\frac{dr_p}{dp}\right) \tag{2.8-12}$$

or

$$\frac{dr_p}{dp} = \zeta_p D \frac{\partial \alpha}{\partial p} \tag{2.8-13}$$

where ζ_p is the defocusing factor

$$\zeta_p^{-1} = 1 - D\frac{\partial \alpha}{\partial r_p} = 1 - D\left(\frac{d\alpha^A}{dr_p} + p\frac{dI}{dr_p}\right) \tag{2.8-14}$$

The differential expression in Eq. (2.8-13) should be valid provided $\left|\zeta^{-1}\right| > 0$, but its validity will progressively worsen as the ray nears a caustic surface where $\zeta^{-1} \to 0$. Combining Eqs. (2.8-11) and (2.8-13) yields

$$\frac{d\alpha}{dp} = \zeta_p \frac{\partial \alpha}{\partial p} = \zeta_p \frac{\partial \alpha^I}{\partial p} = \zeta_p I(r_p) \tag{2.8-15}$$

Let $\bar{r}$ be the position of the tangent point of the L1 and L2 rays in the absence of an ionosphere, or, equivalently, when $p = 0$. We expand $\alpha = \alpha\left[p, r_p\right]$ about this position to obtain

$$\alpha_p = \alpha\left[p, r_p\right] = \alpha[0, \bar{r}] + p\left(\frac{d\alpha_p}{dp}\right)_{\bar{r}} + \frac{1}{2}p^2\left(\frac{d^2\alpha_p}{dp^2}\right)_{\bar{r}} + \cdots \tag{2.8-16}$$

where $\bar{r}$ is given to first order in p by

$$\bar{r} = \frac{p_1 r_{p_2} - p_2 r_{p_1}}{p_1 - p_2} = \frac{f_1^2 r_{p_1} - f_2^2 r_{p_2}}{f_1^2 - f_2^2} \doteq 2.546 r_{p_1} - 1.546 r_{p_2} \tag{2.8-17}$$

Evaluating Eq. (2.8-16) at p_1 and p_2 and linearly combining to eliminate the first-order term in p, we obtain

$$\alpha_3 = \frac{p_1 \alpha_{p_2} - p_2 \alpha_{p_1}}{p_1 - p_2} = \alpha[0, \bar{r}] - \frac{1}{2} p_1 p_2 \left(\frac{d^2 \alpha_p}{dp^2} \right)_{\bar{r}} + \cdots \tag{2.8-18}$$

or, upon using Eq. (2.8-15),

$$\alpha_3 \doteq \alpha[0, \bar{r}] - \frac{1}{2} D \left[\zeta_{p_1} \zeta_{p_2} \frac{\partial}{\partial r_p} \left(\alpha_{p_1}^I \alpha_{p_2}^I \right) \right]_{\bar{r}} \tag{2.8-19}$$

Figure 2-21 shows this linear combination of bending angles using the excess L1 and L2 Doppler data from the occultation shown in Fig. 2-16. At these altitudes (~100 km), α_3 should be zero. It is not quite zero because of ray splitting and the presence of other higher-order effects. Inasmuch as the magnitude of the bending angle itself from the ionosphere is only 0.1 to 0.2 mrad, the error in the α_3 linear combination is as large as the correction itself, which suggests that an alternative technique for correcting for ray splitting and higher-order effects might be advisable [19].

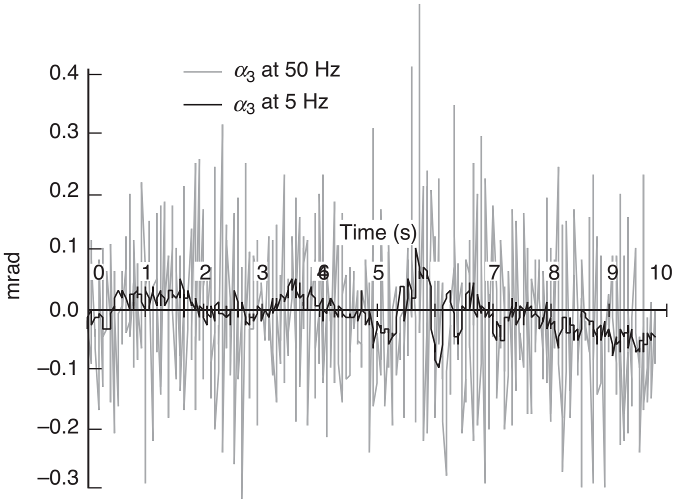

Fig. 2-21. Ionosphere-free bending angle α_3 for the same occultation and epoch shown in Figs. 2-15, 2-16, and 2-20.

The ray-splitting term involves the gradient of the square of the ionospheric bending, which is a consequence of the stationary-phase condition on the ray. When $\bar{r}$ lies above the atmosphere, $\alpha[0,\bar{r}] \equiv 0$, and for this case we have

$$\left.\begin{aligned} &\alpha_3 \doteq -Dp_1p_2\left[\frac{I(r)I'(r)}{\left(1-Dp_1I'(r)\right)\left(1-Dp_2I'(r)\right)}\right]_{r=\bar{r}} \\ &\alpha^A(\bar{r}) \equiv 0 \end{aligned}\right\} \tag{2.8-20}$$

Thus, in the ionosphere the linear combination of the observables α_{p_1} and α_{p_2} that forms α_3 also provides a measure of ray-splitting effects. For the case of spherical symmetry and where n_e denotes the local transient in electron density, the integral $I'\left(r_p\right)$ is given by

$$I'\left(r_p\right) \doteq 2\kappa r_p \int_{r_p}^{\infty} \frac{\sqrt{r^2-r_p^2}}{r^2} n_e' dr + 2\kappa r_p^2 \int_{r_p}^{\infty} \frac{n_e''}{r\sqrt{r^2-r_p^2}} dr \doteq 2\kappa r_p \int_{r_p}^{\infty} \frac{n_e''}{\sqrt{r^2-r_p^2}} dr \tag{2.8-21}$$

which normally is dominated by n_e'' for r near r_p when r_p lies in the ionosphere. Also, from Eq. (2.8-16) we can obtain determinations of $I\left(r_p\right)$ and $I'\left(r_p\right)$ from the observables α_{p_1} and α_{p_2} when $\bar{r}$ lies in the ionosphere. One obtains

$$I(\bar{r}) = \frac{\alpha_{p_1}\alpha_{p_2}\left(p_2-p_1\right)}{p_1p_2\left(\alpha_{p_2}-\alpha_{p_1}\right)}, \ \alpha[0,\bar{r}] \equiv 0 \tag{2.8-22}$$

and

$$I'(\bar{r}) = \frac{1}{D}\frac{\alpha_{p_2}p_1-\alpha_{p_1}p_2}{p_1p_2\left(\alpha_{p_2}-\alpha_{p_1}\right)}, \ \alpha[0,\bar{r}] \equiv 0 \tag{2.8-23}$$

We note that the ray splitting itself is given by

$$r_{p_2}-r_{p_1} = D\left(\alpha_{p_2}-\alpha_{p_1}\right) \doteq D\left(p_2-p_1\right)\left(\zeta_p I(r)\right)_{\bar{r}} \tag{2.8-24}$$

A suggestion that has been made independently by several investigators is to offset the epochs of the Doppler measurements to eliminate (to first order) the ionosphere effects on derived bending angles—in effect, to let the L1 ray "catch up" with the L2 ray. Let Δt_p be the offset of the observational epoch on the "p_kth" carrier for which the change in derived bending angle at that offset epoch nulls (to first order) the dispersive term in Eq. (2.8-16). That is, we require that $\Delta t_p \, d\alpha_p / dt = -\zeta_p \, \alpha_p^I$. From Eqs. (2.2-5) and (2.2-6), we have

$$\frac{d\alpha_p}{dt} = \dot{r}_b \zeta_p \frac{\partial \alpha_p}{\partial r_p} \tag{2.8-25}$$

and, from Eq. (2.8-16), we obtain by expanding through the first derivative in time

$$\begin{aligned} \alpha_p \Big|_{t=t_o+\Delta t_p} &= \alpha_p \Big|_{t_o} + \Delta\alpha_p = \alpha[0,\bar{r}] + \zeta_p \alpha_p^I \\ &+ \left(\dot{r}_b \zeta_p \frac{\partial}{\partial \bar{r}} \alpha[0,\bar{r}] + \dot{r}_b \zeta_p \frac{d}{d\bar{r}} \left(\zeta_p \alpha_p^I \right) \right) \Delta t_p \end{aligned} \tag{2.8-26}$$

from which it follows that to first order

$$\Delta t_p = -\left[\frac{\alpha_p^I}{\dot{r}_b} \left(\frac{\partial}{\partial r_p} \alpha[0, r_p] \right)^{-1} \right]_{r_p = \bar{r}} \tag{2.8-27}$$

Forming the "α_3" linear combination again, but with an offset epoch applied to each carrier according to Eq. (2.8-27), one obtains, when $D|\partial\alpha/\partial r| << 1$,

$$\begin{aligned} \tilde{\alpha}_3 &= \frac{p_1 \alpha_{p_2} \Big|_{t_o+\Delta t_2} - p_2 \alpha_{p_1} \Big|_{t_o+\Delta t_1}}{p_1 - p_2} = \alpha[0,\bar{r}] - D\Bigg[\left(\zeta_{p_1} \zeta_{p_2} \right)^2 \\ &\times \left\{ \left(\frac{\partial \alpha}{\partial r} \right) \frac{\partial}{\partial r} \left(\alpha_{p_1}^I \alpha_{p_2}^I \right) - \zeta_{p_1} \zeta_{p_2} \alpha_{p_1}^I \alpha_{p_2}^I \frac{\partial^2 \alpha}{\partial r^2} \right\} \left(\frac{\partial \alpha}{\partial r} \right)^{-1} \Bigg]_{r=\bar{r}} + \cdots \end{aligned} \tag{2.8-28}$$

When $D|\partial\alpha/\partial r| >> 1$, for example, when $\bar{r}$ is located low in the troposphere, then the quadratic error term in Eq. (2.8-28) is more complicated, but it still will carry the quartic power of the defocusing factors ζ_{p_1} and ζ_{p_2} as the leading term. This factor will be small, so the error term usually should be small.

2.8.4 Doppler Information at a Boundary

In the following, the subscript "p_k" is replaced by "$\mathrm{L}k$", $k = 1, 2$. Returning to the problem of determining Δn_e across a boundary, two alternate approaches can be followed. One is to de-trend the L1 and L2 Doppler profiles using the POD-provided profile of ρ, where ρ is the LEO–to–GPS satellite range; the bending angles $\alpha_{\mathrm{L}1}$ and $\alpha_{\mathrm{L}2}$ follow directly from the excess Doppler. This is the approach that was used in [19] to obtain the results shown in Fig. 2-14. The changes in $\alpha_{\mathrm{L}1}$ and $\alpha_{\mathrm{L}2}$ across the boundary form an Abel transform with Δn_e if local spherical symmetry is assumed, which might be a good assumption given the limited altitude range over which Δn_e applies ($\Delta h = \sim 1\,\mathrm{km}$ in Fig. 2-15, and an equivalent horizontal span of $2(\Delta h r_o)^{1/2} = \sim 200\,\mathrm{km}$). Applying the Abel transform to these differenced observable streams would yield Δn_e; however, if the transient is particularly sharp compared to the size of the first Fresnel zone, then we would expect imperfect recovery due to the limitations in geometric optics.

The other approach is simply to use the profile of L1–L2, which does not require POD information. For this approach, we have

$$\left[\alpha_{\mathrm{L}1}(r_{\mathrm{L}1}) - \alpha_{\mathrm{L}2}(r_{\mathrm{L}2})\right] = -\frac{c}{\dot{r}_{\mathrm{LG}}}[\dot{\mathrm{L}}1 - \dot{\mathrm{L}}2] + \mathrm{O}\left[\alpha^2\right] \tag{2.8-29}$$

When spherical symmetry applies, $\Delta\left[\alpha_{\mathrm{L}1}(r) - \alpha_{\mathrm{L}2}(r)\right]$ and $\Delta\left[N_{\mathrm{L}1}(r) - N_{\mathrm{L}2}(r)\right]$ form an Abel transform pair. However, the quantity [L1–L2] in Fig. 2-15 is formed from the measured phases of the two carriers at a common epoch and is so formed to eliminate the geometric range term between the LEO and the occulted GPS satellite. As Eq. (2.8-29) shows, [L1–L2] is affected by ray splitting, and it should be corrected if $\left[r_{L1} - r_{L2}\right]$ significantly differs from zero. Following the earlier discussion and using Eq. (2.8-16), one obtains for the change in the [L1–L2] profile in the neighborhood of the boundary

$$\begin{gathered}\Delta[\dot{\mathrm{L}}1 - \dot{\mathrm{L}}2] \doteq \\ \left(\frac{\dot{r}_b (p_1 - p_2)\Delta I}{c\left(1 - Dp_1 I'\right)\left(1 - Dp_2 I'\right)()}\right) = \frac{\dot{r}_b}{c}\left[\zeta_{p_1}\zeta_{p_2}\Delta\left[\alpha_{L1} - \alpha_{L2}\right]\right]_{r_p = \bar{r}}\end{gathered} \tag{2.8-30}$$

If $\zeta_p^{-1} = 1 - DpI'$ is near unity, one obtains $\Delta I(\bar{r})$ directly from Eq. (2.8-30); otherwise, Eq. (2.8-30) can be evaluated by iteration, first obtaining the profile for $\left[\alpha_{\mathrm{L}1}(r_{\mathrm{L}1}) - \alpha_{\mathrm{L}2}(r_{\mathrm{L}2})\right]$ from Eq. (2.8-29) and then using this profile over time to obtain an estimate of *I*. Alternatively, one can obtain an estimate of $r_{\mathrm{L}1} - r_{\mathrm{L}2}$ from the difference in epochs of fadeout onset; use of Eq. (2.8-24) and an assumed value for D yields an estimate of *I*.

Finally, correcting Eq. (2.8-29) for ray splitting and applying an Abel transform (when spherical symmetry holds) to Eq. (2.8-30) yields Δn_e in the vicinity of r_o:

$$\begin{aligned} \Delta n_e(\bar{r}) &\doteq \frac{f_1 f_2}{\kappa\pi} \frac{f_1 f}{f_1^2 - f_{2\,2}^2} \int_{\bar{r}}^{\infty} \frac{\Delta\left[\alpha_{L1}(r) - \alpha_{L2}(r)\right]}{\sqrt{r^2 - \bar{r}^2}} dr \\ &= 3.03x10^{16} \int_{\bar{r}}^{\infty} \frac{\Delta\left[\alpha_{L1}(r) - \alpha_{L2}(r)\right]}{\sqrt{r^2 - \bar{r}^2}} dr \end{aligned} \tag{2.8-31}$$

With the value for Δn_e obtained from Eq. (2.8-31) and using the L1–L2 profile, one obtains D from Eq. (2.8-5), which is

$$D = r_o \frac{f_{Lk}^2}{2\kappa \Delta n_e(r_o)} \left(\frac{2B_{Lk}}{3r_o} \right) \tag{2.8-32}$$

The value for *D* so obtained is around 2800 km for the occultation in Fig. 2-15.

2.8.5 Fresnel Effects Using an Improved Electron Density Model

One can improve on the Case A model somewhat. Comparing Figs. 2-15 and 2-18, one notes that a simple model given by a discontinuous electron density at a boundary is not sufficient to explain all of the features in Fig. 2-15 in the vicinity of r_o. Prior to onset of the fadeout zone, Fig. 2-15 shows an SNR_V flaring relative to the base SNR_V that is approximately 33 percent in L1 and nearly 50 percent in L2, which are far in excess of the Fresnel overshoot shown in Fig. 2-18. This suggests that another caustic geometry applies just above the boundary, in addition to the caustic below. In this case, it follows that, in addition to the large positive gradient in electron density just below the boundary, the gradient must become negative just above the boundary (to obtain a positive gradient in refractivity in this region and, therefore, a negative increment in bending angle). Moreover, the dip in the excess Doppler profiles shown in Fig. 2-16 just prior to onset of the fadeout followed by an abrupt rise just afterwards also suggests a sharp negative gradient in electron density above the boundary and a positive gradient below. Apparently, the electrodynamic and electrochemical environment in this region can result in a concentration of free electrons confined to a narrow range of altitudes at roughly 100 km just above the recombination zone. This is depicted in Fig. 2-22. A sporadic E-layer roughly fits this description.

We use a simpler model that will generate caustics above and below the boundary as well as the sharp negative and positive changes in excess Doppler. It is given by

$$\Delta n_e = \begin{Bmatrix} \dfrac{\Delta n_e(r_o)}{1+\gamma}\left[1+\gamma \exp\left(-\beta(r-r_o)\right)\right], \ r \geq r_o \\ 0, \ r < r_o \end{Bmatrix} \tag{2.8-33}$$

Following the same methodology as given for Case A (with $\alpha^+ = 0$), three regimes will apply: the region above the boundary, the narrow region between the critical refraction altitude up to the boundary, and the region below the critical refraction altitude. One obtains for $\Delta\alpha(r)$ in the two regimes

$$\Delta\alpha = \begin{Bmatrix} \dfrac{\Delta N_o \gamma}{1+\gamma}\sqrt{2\pi\beta r}\exp\left(-\beta(r-r_o)\right), \ r \geq r_o \\ \dfrac{\Delta N_o \gamma}{1+\gamma}\sqrt{2\pi\beta r}\,\mathrm{F}[\sigma] + \Delta\alpha_S, \ r \leq \tilde{r} < r_o \end{Bmatrix} \tag{2.8-34}$$

where $\Delta\alpha_S$ is the discrete Snell bending-angle term given in Eq. (2.8-2), $\Delta N_o = -\kappa \Delta n_e(r_o)/f^2$, $\mathrm{F}[\sigma]$ is defined by Eq. (2.3-28), $\sigma = \sqrt{\beta(r_o - r)}$, and $\tilde{r} = (1+\Delta N_o) r_o$, which denotes the critical refraction point for $\Delta n_e(r_o) > 0$. The corresponding refractivity and bending-angle profiles are shown in Figs. 2-23 and 2-24. This would produce a dip in excess Doppler prior to onset of fadeout and an abrupt rise just after onset, such as shown in Fig. 2-16.

The free parameters in this model are ΔN_o, γ, and β, and we can adjust them using the relationships developed in this section to best fit the observed profiles in Figs. 2-15 and 2-16. Using the stationary-phase condition for the

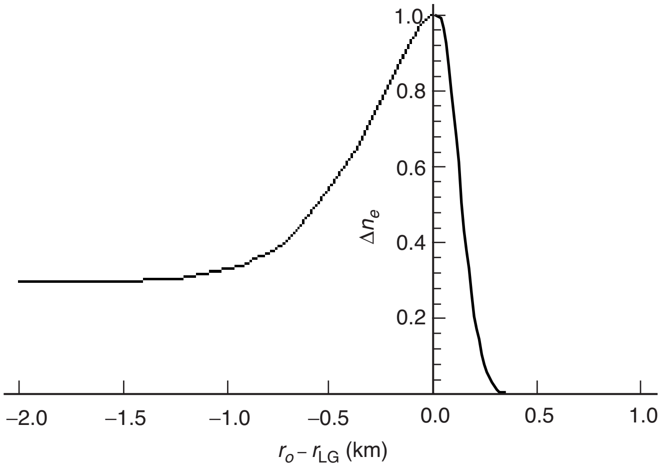

Fig. 2-22. Trial electron density profile near the boundary, $r = r_o$.

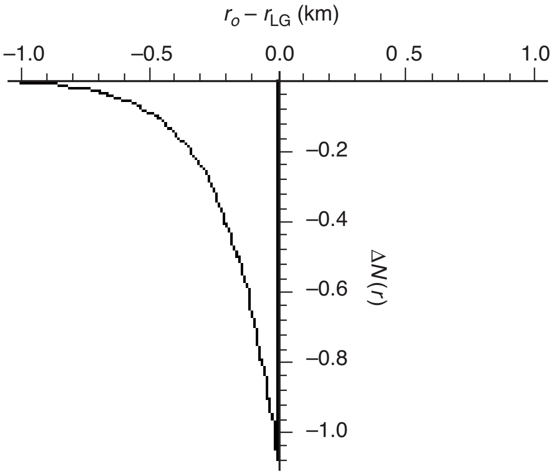

Fig. 2-23. Trial refractivity profile according to Eq. (2.8-33).

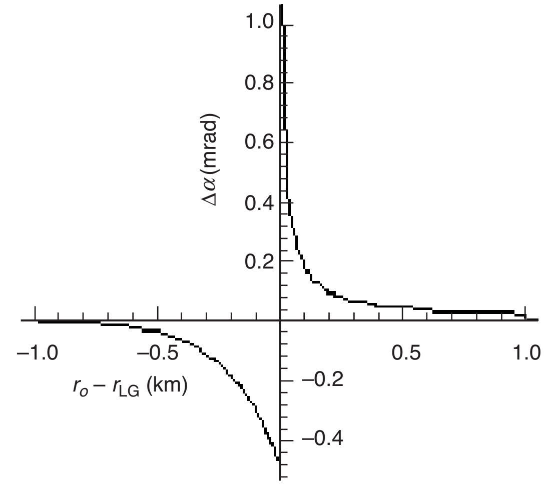

Fig. 2-24. Refractive bending angle from Eq. (2.8-34).

thin-screen model, one integrates Eq. (2.8-34) to obtain the phase delay $\varepsilon(h)$, which is to be embedded in the thin screen. It is given by

$$\varepsilon(h) = \left\{ \begin{array}{l} \dfrac{2\pi}{\lambda\beta}\alpha(r), \quad r > \tilde{r} \\ \dfrac{2\pi}{\lambda\beta}\left[\alpha(r) - \Delta\alpha_S(r) + 2\sigma\right] + \varepsilon_S(r), \quad r \le \tilde{r} < r_o \end{array} \right\} \tag{2.8-35}$$

where $\Delta\alpha_S$ and ε_S are the Snell terms given by Eqs. (2.8-2) and (2.8-3), respectively, and $h = r - R$. From the Rayleigh–Sommerfeld integration in Eq. (2.5-1), one obtains the observed Fresnel response to this kind of electron density distribution.

Figure 2-25 shows the Fresnel perturbations in observed signal amplitude in the vicinity of r_o for $\Delta N_o = -1.2\times10^{-6}$, $\gamma = 10$, and $\beta = 5\,\text{km}^{-1}$. The similarity of Fig. 2-25 in the fadeout zone, including the neighboring crests due to flaring, to those in Fig. 2-15 should be noted. Figure 2-26 shows the excess Doppler, including the Fresnel perturbations that result from the assumed electron density model given in Eq. (2.8-33). The black curve is a smoothed version obtained by applying a 1/4-s-wide box filter to the 40-Hz samples. The qualitative agreement between Figs. 2-16 and 2-26 should be noted in the vicinity of $r_{\text{LG}} = r_o$. The abscissa range shown in Figs. 2-25 and 2-26 is equivalent to about 3 s of observations.

Fresnel expressions for estimating minimum signal amplitude in a trough or maximum signal amplitude at the crest of a flaring region, and for calculating the positions of these features in h_{LG}-space, will be found in Appendix D. However, from geometric optics we can estimate the altitude of the flaring using the condition that $dr_{\text{LG}} / dr = 0$ at the first contact point with a caustic. Using the thin-screen relationship in Eqs. (2.2-5) and (2.8-34) for $\Delta\alpha$, one obtains for the position $r_{\text{LG}}^{\dagger}$ of the first contact point at the upper caustic

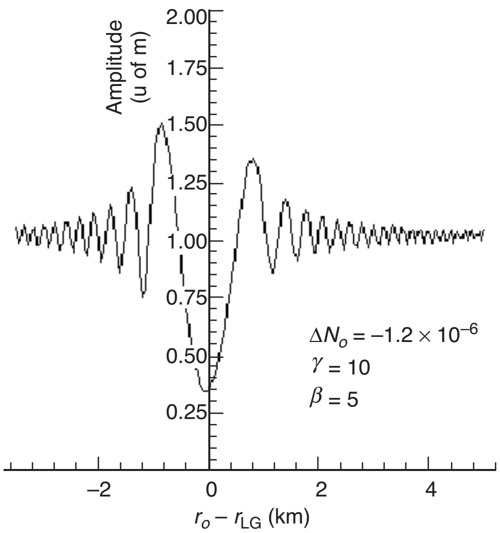

Fig. 2-25. Fresnel perturbation in signal amplitude using the thin-screen phase profile from Eq. (2.8-35). This figure should be compared with Fig. 2-15.

$$r_{\mathrm{LG}}^{\dagger}=r_o+r^{\dagger}+\beta^{-1}=r_o+\frac{1}{\beta}\left(\log\left[-\Delta N_o D\gamma\frac{\left(2\pi r_o\beta^3\right)^{1/2}}{1+\gamma}\right]+1\right) \qquad (2.8\text{-}36)$$

which predicts the contact point (where flaring should be near maximum) at $r_o - r_{\mathrm{LG}}^{\dagger} = -0.6\,\mathrm{km}$, or about 0.2 s prior to the time of impact of the ray with the boundary. The exact location of the point of maximum amplitude in Fig. 2-25, which can be accurately predicted using the third-order theory in Appendix D, is $r_o - r_{\mathrm{LG}}^{\dagger} = -0.84\,\mathrm{km}$.

Finally, we can apply the Abel transform to the L1 Doppler series shown in Fig. 2-26, which is based on the thin-screen model using the refractivity model given in Eq. (2.8-33) and shown in Fig. 2-23. This will give us one qualitative estimate of the fidelity of the Abel transform for two transient situations: for $r_{\mathrm{LG}} > r_o$ the transition scale in refractivity is comparable to the scale of the first Fresnel zone, but at $r_{\mathrm{LG}} = r_o$ there is a sharp transition, a discontinuity $\Delta N_o = -1.2\times10^{-6}$, in this case. The recovery of the refractivity profile based on the Abel transform and the original model are shown in Fig. 2-27, as too is the truth model $N(r)$ given by Eq. (2.8-33). Here $N(-\infty) = \Delta N_o/(1+\gamma)$. Even the smoothed profile (at 4 Hz) shows a bias as a result of the sharp transition in true refractivity at the boundary.

2.9 The Error in the Recovered Refractivity Resulting from Fresnel Phase Perturbations

Since the Abel transform is a linear operator, we can use it, if spherical symmetry applies, to assess the error in the recovered refractivity that results

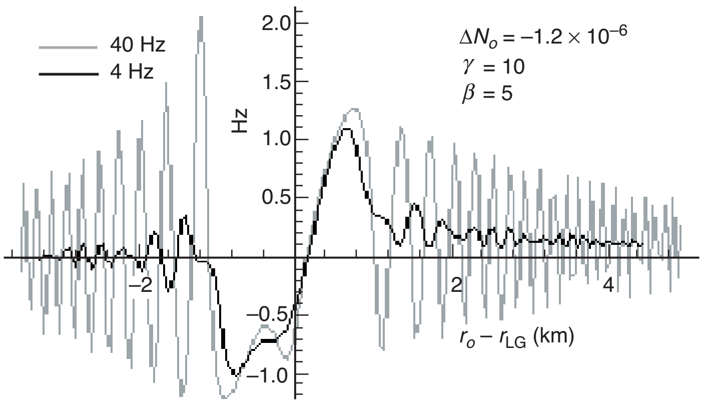

Fig. 2-26. Residual excess L1 Doppler including Fresnel perturbations for the electron density model given in Eq. (2.8-33). This figure should be compared with Fig. 2-16.

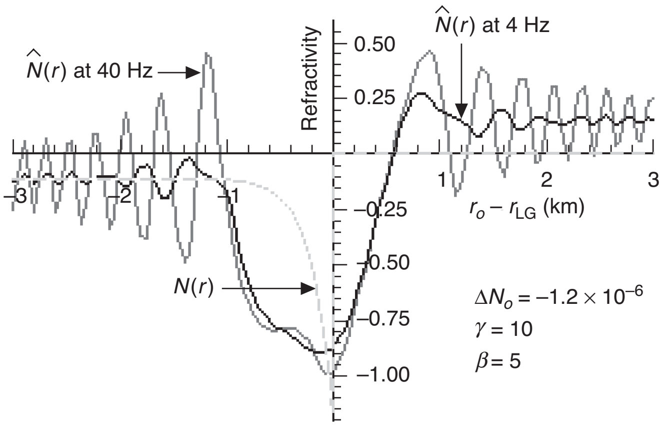

Fig. 2-27. Recovered refractivity profile $\hat{N}(r)$ obtained from the Abel transform versus the truth model $N(r)$ given by Eq. (2.8-33).

from using a geometric optics approach when Fresnel phase perturbations are present (see Appendix F). Let $\delta\hat{N}$ ($\times 10^{-6}$) be the error in recovered refractivity, which is obtained from the Abel transform by the expression

$$\delta\hat{N}(a) = \frac{1}{\pi}\int_a^\infty \frac{\delta\alpha(\mu)}{\sqrt{\mu^2 - a^2}} d\mu \tag{2.9-1}$$

where a is the impact parameter for the ray path. For error assessment purposes, it can be taken as $R_E + h_o$, the altitude in the thin screen corresponding to the real altitude of the tangency point of the ray to the Earth's limb, plus the Earth's radius R_E. Now the error in the inferred bending angle is given to a good approximation by

$$\delta\alpha = \frac{\lambda}{2\pi}\frac{\delta\dot{\varphi}}{\dot{h}_{\mathrm{LG}}} \doteq \frac{\lambda}{2\pi}\frac{d(\delta\varphi)}{dh_{\mathrm{LG}}} \tag{2.9-2}$$

where $\delta\varphi$ is the phase perturbation caused by some error source. Let $\delta\varphi_F\left(h_{\mathrm{LG}}\right)$ be the phase perturbation due to Fresnel effects, such as that shown in Fig. 2-10. Then Eq. (2.9-1) becomes

$$\delta N_F(h) = \frac{\lambda}{2\pi^2}\int_h^\infty \frac{d\left(\delta\varphi_F\right)}{dh_{\mathrm{LG}}} \frac{1}{\sqrt{\mu^2 - r^2}} d\mu \tag{2.9-3}$$

where the dummy variable μ is given by $\mu = R_E + h'$ and r is given by $r = R_E + h$. To avoid differentiating $\delta\varphi_F(h_b)$, Eq. (2.9-3) may be written in an alternate form:

$$-\frac{\lambda}{2r\pi^2}\frac{d}{dr}\int_h^\infty \left\{\delta\varphi_F\left[h_{\mathrm{LG}}(h')\right]\left[\frac{d\zeta}{dh'}\sqrt{\mu^2-r^2}+\frac{\mu\zeta(h')}{\sqrt{\mu^2-r^2}}\right]\right\}d\mu \qquad (2.9\text{-}4)$$

where ζ is the defocusing factor, $\zeta = dh/dh_{\mathrm{LG}}$.

For an atmosphere in local thermodynamic equilibrium and of known constituents, the consequent error in the retrieved temperature profile due to unmodeled Fresnel perturbations follows from the relation

$$\left.\frac{\delta T}{T}\right|_h = -\left.\frac{\delta N}{N}\right|_h + \frac{\int_h^\infty \delta N dh}{\int_h^\infty N dh} \qquad (2.9\text{-}5)$$

When the spectrum of $\delta N(h)$ has most of its power at short wavelengths (i.e., <1 km), which is the case for Fresnel-induced variations, then the first term on the RHS of Eq. (2.9-5) will be dominant. In any case, the second term is given by

$$\int_h^\infty \delta N_F(h')dh' = \frac{\lambda}{2\pi^2}\int_h^\infty \frac{d(\delta\varphi_F)}{dh_{\mathrm{LG}}}\cos^{-1}\left(\frac{r}{\mu}\right)d\mu,\ \ \mu = R_E + h' \qquad (2.9\text{-}6)$$

Figures 2-28(a) through 2-28(d) provide an evaluation of the Fresnel effects on the recovery of $N(h)$ and $T(h)$ for the Case C example given in Fig. 2-10(b), with a lapse rate discontinuity of 14 K/km. Here, as in Fig. 2-10(b), the geometric optics component due to the lapse rate discontinuity has been suppressed because the Abel transform would recover this component. A lapse rate discontinuity of 14 K/km results in an error of a few tenths of a kelvin in retrieved temperature in the vicinity of the discontinuity. As expected, the maximum error occurs near first contact with the caustic, which for the + ray (in Figs. 2-28(c) and 2-28(d) the abscissa is the turning point altitude h^+ of the + ray) is located at $h^+(2) - h_o = +0.16\,\mathrm{km}$, that is, it is *above* the discontinuity. The – rays are located at $h^{a,b}(2) - h_o = -0.16\,\mathrm{km}$. This example clearly shows that even though the altitude differences of the turning points of these multipath rays are well within the first Fresnel zone, a Fresnel diffraction treatment is required to properly account for the effects of the caustic. On the other hand, SNR limitations, at least on GPS/MET, constrain the utility of high-resolution Doppler sampling. A typical averaging interval used by investigators

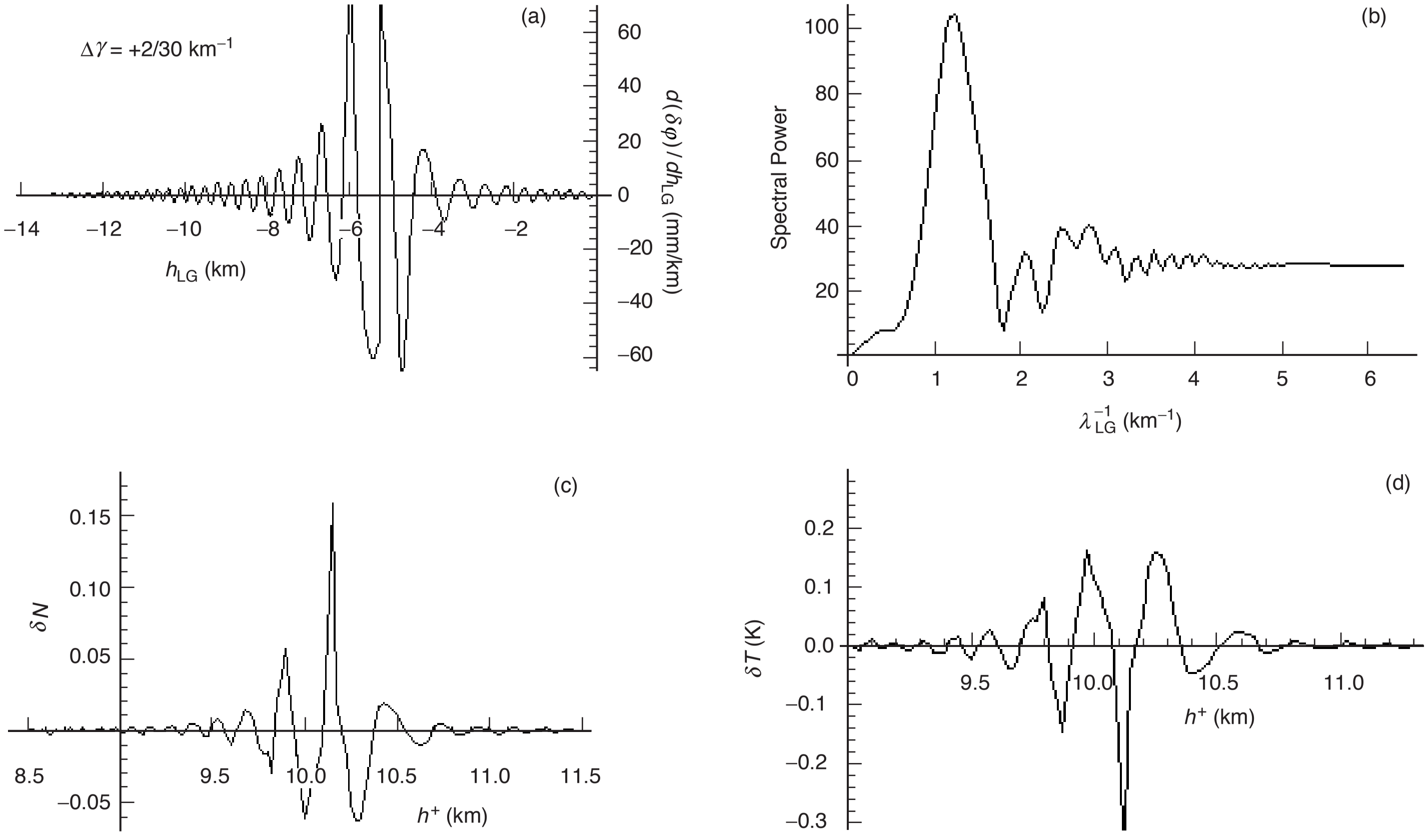

Fig. 2-28. Fresnel perturbations and errors for Case C with a lapse rate discontinuity of 14 K/km at an altitude of 10 km: (a) Doppler perturbations, (b) Fourier spectrum of the Doppler perturbation, (c) refractivity error, and (d) temperature error.

in analyzing GPS/MET data has been about 1/3 s, or about 1 km in h_{LG}, which largely washes out Fresnel effects [see Figs. 2-28(c) and 2-28(d)].

2.10 Fresnel Transform Techniques

Because of their highly localized character, transients such as those observed across a sporadic E-layer would appear to be good candidates for treating by inverse diffraction techniques to secure sub-Fresnel vertical resolutions. Limitations in resolution using inverse diffraction techniques can arise from along-track inhomogeneity, from an uncertainty in the cross-track gradient of the TEC (total along-track electron content per unit columnar area, 1 TEC = 10^{16} electrons/m^2) between the LEO and the layer, and also from the uncertainty in the effective or mean along-track distance of the layer from the LEO. Any model error that causes a loss of coherence as the integration limits are broadened between the integrands ($E(h_{\mathrm{LG}})\exp[i\Psi(h_{\mathrm{LG}})]$ in the diffraction integral [Eq. (2.5-1)] and $A(h)\exp[i\Phi(h,h_{\mathrm{LG}})]$) in its Fresnel transform will degrade the resolution.

For a spherical shell with a 1-km vertical thickness, the along-track extent of the layer would be roughly 200 km. It can be shown [4,5] that δh, the vertical resolution limit due to an error δD in the *assumed* along-track distance, is given by

$$\delta h \geq 1.25\sqrt{\frac{\lambda D}{2}\frac{\delta D}{D}} \tag{2.10-1}$$

Thus, an uncertainty of 5 percent in the along-track distance (about 150 km) translates into a lower resolution limit of about 20 percent of the radius of the first Fresnel zone; the limit scales as the square root of the error in the knowledge of the distance D of the layer.

Uncertainty in the cross-track gradient in TEC between the LEO and the layer averaged over the vertical extent associated with the diffraction integration interval (a few kilometers) is another problem. Although a time sequence of dual-band phase measurements allows a determination of the overall (LEO-to-GPS) cross-track gradient in TEC, the component of cross-track TEC gradient arising from the electron density spatial variability *between* the LEO and the boundary layer is partially unknown. This component can "leak into" the diffraction and thereby corrupt the resolution because of the 13 percent difference in Fresnel zone radii between the L1 and L2 carriers. It can be shown [4] that an error in the knowledge of this cross-track gradient degrades the resolution by

$$\delta h \geq D \frac{1.75 \times 10^{18}}{f^2} \delta\left(\frac{d(TEC)}{dh_{\mathrm{LG}}} \right) \text{ m} \tag{2.10-2}$$

It follows that a near-field error of 0.1 TEC/km maps into a vertical resolution lower limit of about 30 percent of the radius of the first Fresnel zone; note that this limit scales linearly with the error in the near-field component of cross-track TEC gradient.

With these caveats in mind[7] (and others; see, for example, [5]), we briefly discuss enhanced resolution using the Fresnel transform. The phase perturbation in the thin screen due to $\delta\alpha$ theoretically can be recovered with enhanced resolution from the observed phase and amplitude variations by using Fresnel transform techniques. We may generalize Eq. (2.5-1) to the form

$$\left.\begin{aligned} &E(u_{\mathrm{LG}})\exp\big(i\psi(u_{\mathrm{LG}})\big) = \frac{1}{1+i}\int_{-\infty}^{\infty} A(u)\exp\left(i\left(\varphi(u) + \frac{\pi}{2}(u - u_{\mathrm{LG}})^2 \right)\right) du \\ &u = h\left(\frac{2}{\lambda D} \right)^{\frac{1}{2}}, \quad u_{\mathrm{LG}} = h_{\mathrm{LG}}\left(\frac{2}{\lambda D} \right)^{1/2} \end{aligned}\right\} \tag{2.10-3}$$

Here u is the thin-screen altitude h expressed in Fresnel units, and similarly for u_{LG}, $\varphi(u)$ is the thin-screen phase profile, and $A(u)$ is the normalized amplitude of the radiation emitted by the thin screen ($A(u)$ would be unity throughout the screen when no attenuation exists). Then $A(u)\exp[i\varphi(u)]$ and $E(u_{\mathrm{LG}})\exp\left[i\psi(u_{\mathrm{LG}})\right]$ form a Fresnel transform pair. Multiplying the integral in Eq. (2.10-3) by $\exp\left[-i(v-u_{\mathrm{LG}})^2\right]$, integrating on u_{LG}, and using the Dirac delta function (see Appendix A) yields

$$A(u)\exp\big(i\varphi(u)\big) = \frac{1}{1-i}\int_{-\infty}^{\infty} E(u_{\mathrm{LG}})\exp\left(i\left(\psi(u_{\mathrm{LG}}) - \frac{\pi}{2}(u - u_{\mathrm{LG}})^2 \right)\right) du_{\mathrm{LG}} \tag{2.10-4}$$

The perturbation in atmospheric refractivity would be obtained from

$$\delta N(h) \doteq 10^6 \frac{\lambda}{2\pi^2 a} \frac{d}{da} \int_a^{\infty} \frac{\xi \varepsilon(h)}{\sqrt{\xi^2 - a^2}} d\xi \tag{2.10-5}$$

---

[7] The effect of along-track inhomogeneities on resolution can be assessed using the multiple-screen approach [10,11] combined with a two-dimensional random walk model to account for variations in refractivity in the along-track and radial directions.

where $\xi = h + R_E$, $h = a\sec\alpha - R_E$. Here $\varepsilon(h)$ is the perturbation in the thin-screen phase profile obtained by subtracting the reference profile $\varphi_{\mathrm{Ref}}(h)$ from the profile $\hat{\varphi}(h)$ recovered from the inverse Fresnel transform in Eq. (2.10-4).

Strategies for using Fresnel transform techniques as outlined by Eqs. (2.10-4) and (2.10-5) to achieve resolutions past the Fresnel limit are discussed in [4,5]. Sub-Fresnel-zone resolution can be obtained in certain physical situations. This was clearly demonstrated by [5] in recovering certain physical parameters of Saturn's rings from radio occultation observations. The extent to which these techniques are applicable to the Earth's atmosphere and ionosphere will depend in part on how laminar the atmosphere is and how benign the ionosphere is.

2.10.1 Adjoining the Constraint $A(u) \equiv 1$

One could strengthen the recovery of the thin-screen phase profile by adjoining a constraining condition to Eq. (2.10-3) requiring that $A(u) \equiv 1$ and applying some least-squares criterion to the resulting linear operators. One of the shortcomings of the formal application of the inverse Fresnel transform in Eq. (2.10-4) is that this approach does not take into account the presence of errors in the phase and amplitude observations and one is free to adjust the recovered values of $A(u)$ accordingly without a priori information about $A(u)$ being imposed.

A straightforward alternative approach is to linearize the forward Fresnel transform in Eq. (2.10-3) in terms of the variables φ, E, and ψ by varying them about their nominal values and then to convert the resulting variational equations, both the in-phase and quadrature components, into a discrete, over-determined, linear least-squares estimation system. The state vector is the thin-screen phase sequence $\{\delta\varphi\} = \{\varepsilon\}$. These vector elements are to be adjusted and iterated according to some least-squares criterion that minimizes the weighted root-mean-square (rms) of the "O–C" vector residuals $\{\delta E\}$ and $\{\delta\psi\}$. It can be shown with $A(u) \equiv 1$ that the linearized version of Eq. (2.10-3) is given by

$$\left.\begin{array}{r}
\begin{Bmatrix} \delta E(u_{\mathrm{LG}}) \\ E(u_{\mathrm{LG}})\delta\psi(u_{\mathrm{LG}}) \end{Bmatrix} = \dfrac{1}{2}\displaystyle\int_{-\infty}^{\infty} \begin{Bmatrix} \cos\Omega + \sin\Omega \\ \cos\Omega - \sin\Omega \end{Bmatrix} \varepsilon(u)du \\
\Omega(u, u_{\mathrm{LG}}) = \psi(u_{\mathrm{LG}}) - \varphi(u) - \pi(u - u_{\mathrm{LG}})^2/2
\end{array}\right\} \qquad (2.10\text{-}6)$$

Converting the integrals in Eq. (2.10-6) into discrete forms to handle discrete sampling should be done with caution because it can introduce subtle effects related to aliasing, sampling bandwidth versus the spectral breadth of the phase and amplitude observations, and SNR levels.

2.10.2 Diffraction Integral with Multipath

One can generalize Eqs. (2.10-4) and (2.10-5) further to account for multipath, following the lines suggested in Chapter 3, footnote 2. Suppose the LEO GPS receiver has successfully "tracked" the amplitude and phase profiles of the individual tones, or has at least operated in an open-loop mode so as to enable this recovery using Fourier transform techniques. Let $E^{(j)}(u_{\mathrm{LG}})$ and $\psi^{(j)}(u_{\mathrm{LG}})$ be the amplitude and phase of the j th tone "tracked" by the receiver on the LEO. Then the stationary-phase condition must apply to each tone, and it follows that

$$E^{(j)}(u_{\mathrm{LG}})e^{i\psi^{(j)}(u_{LG})}=\frac{1}{1+i}\int_{-\infty}^{\infty}A^{(j)}(u)e^{i\left(\varphi^{(j)}(u)+\frac{\pi}{2}(u-u_{\mathrm{LG}})^2\right)}du,\quad j=1,2,\cdots \tag{2.10-7}$$

In this case, $A^{(j)}$ would not be unity because the incident signal power would have to be shared among the tones so that $\sum A^{(j)}=1$. If geometric optics is applicable, the local defocusing factor for each tone could be used to obtain an a priori estimate of the relative values of $A^{(j)}$. The Fresnel transform of Eq. (2.10-7) is given by

$$A^{(j)}(u)e^{i\varphi^{(j)}(u)}=\frac{1}{1-i}\int_{-\infty}^{\infty}E^{(j)}(u_{\mathrm{LG}})e^{i\left(\psi^{(j)}(u_{LG})-\frac{\pi}{2}(u-u_{\mathrm{LG}})^2\right)}du_{\mathrm{LG}},\quad j=1,2,\cdots \tag{2.10-8}$$

2.10.3 A Numerical Example

With the 50-Hz L1 $\mathrm{SNR_V}$ and phase data from the occultation profile shown in Figs. 2-15 and 2-16 as input into Eq. (2.10-4), one can generate estimates of $\varepsilon(u)$ for the layer. We assume that over the few seconds of interest here the phase contribution from the ionosphere at large is essentially constant. One can improve on this procedure by modeling the contribution from the ionosphere at large over this short time interval; for example, it is easily shown that adding a constant Doppler term Δf_d to the excess Doppler series in Fig. 2-16 (to null its mean value) results in an offset in the recovered profile of $\varepsilon(u)$ [from Eq. (2.10-4)] in h-space that is equal to $D\Delta\alpha^I$, or about 100 m; $\Delta\alpha^I$ is the bending-angle increment resulting from the Doppler offset Δf_d. More specifically, we perturb the observed phase: $\psi\to\psi^T+2\pi\Delta f_d t$, where t is a linear function of h_{LG} and $t=0$ when $h_{\mathrm{LG}}=0$. Let $\tilde{A}(u)\mathrm{Exp}[i\tilde{\varphi}(u)]$ be the recovered thin-screen amplitude and phase including the effect of this linear perturbation in observed phase. Then from Eq. (2.10-4) it can be shown that

$$\left.\begin{aligned}
&\tilde{A}(u)e^{i\tilde{\varphi}(u)} = e^{i\frac{\pi}{2}\left((u+\Delta u)^2-u^2\right)}\frac{1}{1-i}\int_{-\infty}^{\infty} E(u_{\mathrm{LG}})e^{i\left(\Psi^T(u_{\mathrm{LG}})-\frac{\pi}{2}(\Delta u+u-u_{\mathrm{LG}})^2\right)}du_{\mathrm{LG}},\\
&\Delta u=\sqrt{\frac{2D}{\lambda}}\Delta\alpha^I,\ \ \Delta\alpha^I=\lambda\frac{\Delta f_d}{\left|\dot{r}_{\mathrm{LG}}\right|},\ \ u=h\left(\frac{2}{\lambda D}\right)^{1/2},\ \ u_{\mathrm{LG}}=h_{\mathrm{LG}}\left(\frac{2}{\lambda D}\right)^{1/2}
\end{aligned}\right\} \tag{2.10-9}$$

Thus, the recovered profiles of $\tilde{A}(u)$ and $\tilde{\varphi}(u)$ are offset[8] in the thin screen by Δu. We neglect this offset Δu here because we are interested in the resolution potential of Eq. (2.10-4).

We renormalize the excess L1 Doppler series shown in Fig. 2-16 so that its average value for the first 5 seconds in this figure is zero. Then we integrate this Doppler series with time to obtain a renormalized L1 phase profile (see Fig. 2-29), which along with the normalized $\mathrm{SNR_V}$ series (its average value over the first 5 seconds is set to unity) are used in Eq. (2.10-4) to obtain the profiles for the thin-screen amplitude and phase in the vicinity of the transient, which from Fig. 2-16 is located around $t = 5.7\,\mathrm{s}$.

As discussed earlier, the integration limits in Eq. (2.10-4) affect the resolution that can be achieved when modeling errors are present. Thus, the optimal integration limit in Eq. (2.10-4) involves a trade-off between accuracy and resolution. For an uncertainty in D of 5 percent, the single-sided integration limit giving the best resolution is about 3 to 4 Fresnel radii, or about 1 to 2 s of elapsed time; the best resolution itself would be given by Eq. (2.10-1), and would be about 20 percent of a Fresnel radius in this case.

---

[8] In Fresnel transform theory, the pair (u_{LG}, u) form a conjugate variable set that is analogous to that formed by spectral number and coordinate position, or by frequency and time in Fourier transform theory, or, equivalently, by position and momentum, or by energy and time in quantum mechanics. The reader can verify that there is an analogous "Heisenberg uncertainty" principle that holds between Δu_{LG} and Δu. In diffraction integral processes, this translates into the following proposition. Suppose that a transient occurs in the thin-screen phase profile of characteristic width Δu (for example, a Gaussian-shaped transient with a 1-σ half-width of Δu). This transient causes a corresponding transient in the observations of characteristic width $\Delta u_{\mathrm{LG}} = \Delta T \dot{r}_{\mathrm{LG}}(2/\lambda D)^{1/2}$ (the envelope of which for this example also would be a Gaussian with a 1-σ half-width of Δu_{LG}). Then the uncertainty principle states that the product $\Delta u \cdot \Delta u_{\mathrm{LG}} \geq 1/\pi$. This inequality is of course a hallmark of classical diffraction: the narrower the aperture through which waves must pass, the broader their overall pattern at reception.

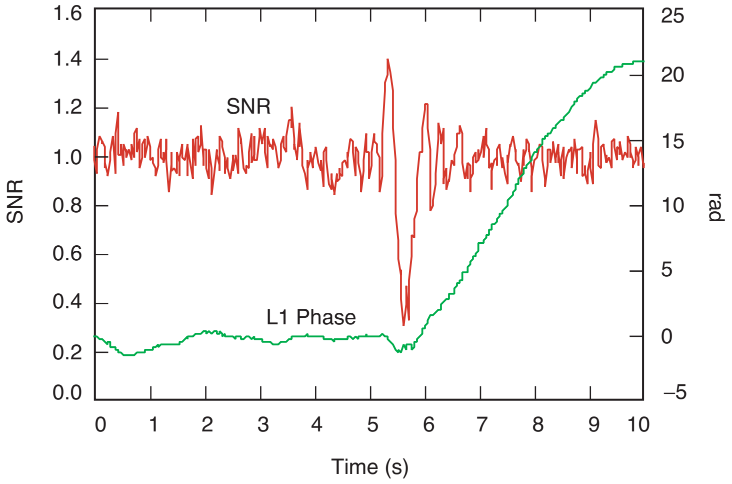

Fig. 2-29. Renormalized L1 SNR and excess phase for the same occultation and epoch shown in Figs. 2-15 and 2-16.

2.10.4 Fresnel Aliasing

The finite temporal sample width (20 ms) of the L1 Doppler series shown in Fig. 2-16 can result in a kind of “Fresnel aliasing” in the convolution; so, one must use integration limits that are well below a certain threshold at which the effects of this aliasing become prominent. Let k be an integer that denotes the position in h_{LG}-space of the LEO–GPS line at the kth observation epoch, and similarly, let j denote a thin-screen position in h-space. When $h_{\mathrm{LG}} = h$, $k = j$; thus, $h_{\mathrm{LG}} = (\dot{r}_{\mathrm{LG}} \Delta T)k + c$ and $h = (\dot{r}_{\mathrm{LG}} \Delta T) j + c$, where c is a constant (we have assumed here that the defocusing is negligible). Here ΔT is the temporal sample interval, 20 ms. In this case, the aliasing threshold for the Fresnel phase in Eq. (2.10-4) is given by the discrete form of the stationary-phase condition

$$\psi(k+1) - B(j-k-1)^2 - \psi(k) + B(j-k)^2 = \pm 2n\pi \tag{2.10-10}$$

where n is an integer and where

$$B = \pi \frac{(\dot{r}_{\mathrm{LG}} \Delta T)^2}{\lambda D} \tag{2.10-11}$$

The principal contributions to the continuous version of the inverse diffraction integral in Eq. (2.10-4) come from the neighborhood(s) around the point(s) where k to the nearest integer satisfies the condition given in Eq. (2.10-10) with $n = 0$. However, in the discrete version of Eq. (2.10-4), pseudo-stationary-phase contributions also arise with $n = \pm 1, \pm 2, \cdots$. Let $M^* = |k^* - j|$ be the aliasing threshold, which must significantly exceed the single-sided range of k values about a given value of j that can be used in the deconvolution to avoid Fresnel aliasing effects. Expanding Eq. (2.10-10) yields

$$M^* = \left| \frac{\lambda D}{\left(\dot{r}_b \Delta T\right)^2} \left(f_d \Delta T \pm n\right) - 1/2 \right| \tag{2.10-12}$$

where f_d is the excess Doppler in hertz. For $n = 1$, M^* is about 240 for this occultation, which exceeds the integration limit for optimal resolution when $\delta D / D \geq 0.3\%$ and when the uncertainty in the near-field cross-track gradient in TEC is greater than ~0. 1 TEC/km.

2.10.5 Numerical Results

The results of this inverse diffraction integral are shown in Figs. 2-29 and 2-30(a) through 2-30(c). Figure 2-30 shows the renormalized excess L1 phase and SNR for the occultation shown in Figs. 2-15 and 2-16. Figure 2-30(a) shows the recovered thin-screen amplitude and phase using Eq. (2.10-4) without applying the a priori constraint $A(u) \equiv 1$. Figure 2-30(b) shows the gradient of the recovered thin-screen phase delay, which is proportional to bending angle. Note the change in time scale in Figs. 2-30(a) through 2-30(c) versus Fig. 2-29. Although the recovered profile for the thin-screen amplitude is roughly unity over most of its time series, it deviates significantly from unity at the transient; it is in this neighborhood where adjoining the a priori constraint might strengthen the recovery of the thin-screen phase profile. On the other hand, it is well known from the mathematical concept of observability in estimation theory that in the presence of mismodeled effects or certain other error sources, one might obtain better least-squares determinations of the parameters of interest by allowing certain other parameters (not well-observed in the first place) a wider latitude in the least-squares process to absorb the effects of such error sources. A good strategy here is not clear without first investigating the spectral nature of the error sources and constructing the sensitivity matrix of the estimated parameter set.

To obtain the electron density profile across the transient, we relate the phase profile in the thin-screen model to the change in refractivity across a thin, spherical ionospheric shell. In thin-screen methodology involving a single screen, the actual along-track phase accumulation through the shell is replaced with a thin-screen phase that is given by the stationary-phase condition through Eq. (2.6-5), which relates the thin-screen phase profile to the refractive bending-angle profile based on geometric optics. For the very small bending angles involved here and assuming local spherical symmetry, we can use the approximate relationship for the thin-screen phase perturbation:

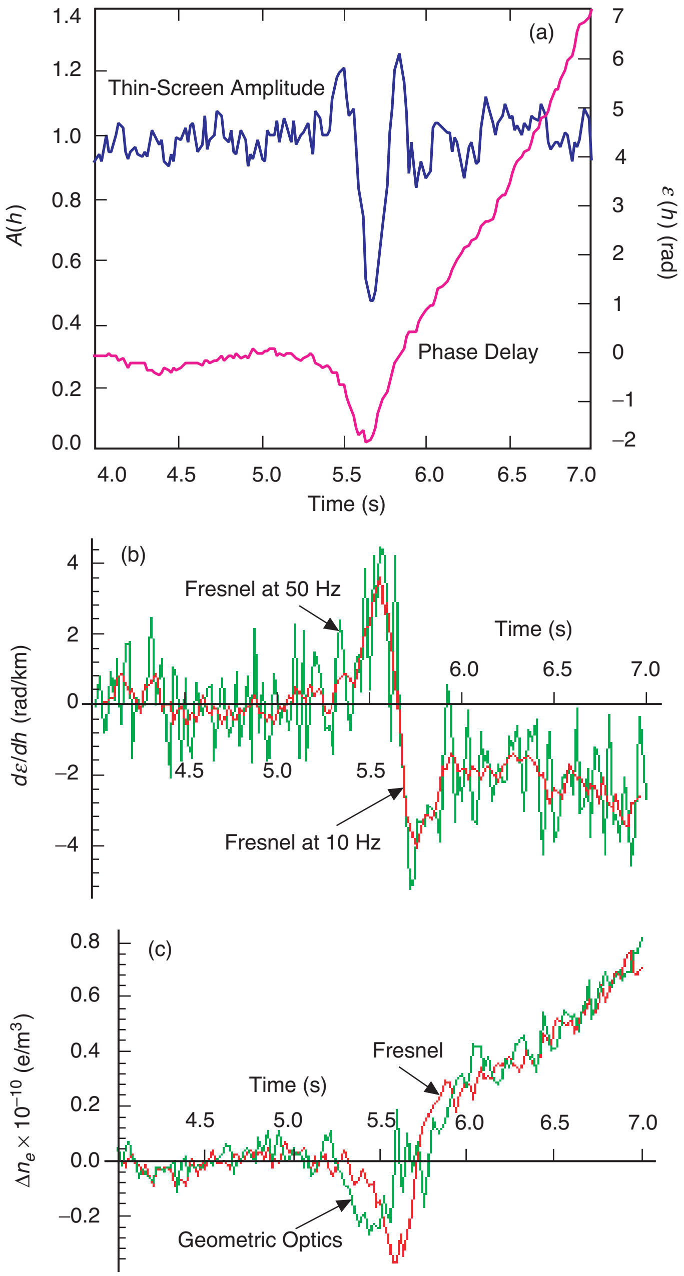

Fig. 2-30. Recovered thin-screen properties from the Fresnel transform without the adjoining a priori constraint $A(u) = 1$: (a) recovered thin-screen phase and amplitude, (b) recovered gradient in phase (bending angle), and (c) change in recovered refractivity using the Fresnel transform versus the Abel transform. The data are the same as in Fig. 2-29. Note the change in time scale.

$$\begin{aligned}\varepsilon(r) &\doteq 2k\int_0^\infty \Delta N(r')ds \\ &= 2k\int_r^\infty \Delta N \frac{r'dr'}{\sqrt{r'^2 - r^2}} = -2k\int_r^\infty \frac{d(\Delta N)}{dr'}\sqrt{r'^2 - r^2}\,dr'\end{aligned} \tag{2.10-13}$$

As Fig. 2-31 shows, even a delta-function-like impulse in refractivity across a thin ionospheric shell results in a lingering effect in the thin-screen phase profile because the rays continue to traverse the shell twice at lower depths. Upon differentiating ε with respect to r and using the Abel transform (and setting $\Delta N(\infty) = 0$), it follows that

$$\Delta N(r) \doteq -\frac{1}{\pi k}\int_r^\infty \frac{d\varepsilon}{d\xi}\frac{d\xi}{\sqrt{\xi^2 - r_o^2}} \tag{2.10-14}$$

which is essentially the same form as given by Eq. (2.9-3). The result of this integration is shown by the curve labeled "Fresnel" in Fig. 2-30(c). The half-width of the bulge in electron density is about 300 m. The curve labeled "Geometric Optics" is the geometric optics version, that is, the profile obtained from applying the Abel transform directly to the L1 Doppler series in Fig. 2-16, renormalized so that the average excess Doppler over the first 5 seconds is zero. As has already been suggested by Fig. 2-27, the geometric optics version "tracks" the medium-to-long wavelength variability (compared to the Fresnel scale) in refractivity fairly well. But just prior to and just after encountering the sharp transition in refractivity at $t = 5.6$ s in Fig. 2-30(c), the fidelity of this algorithm appears to degrade significantly.

We also can compare the constrained and unconstrained versions of the

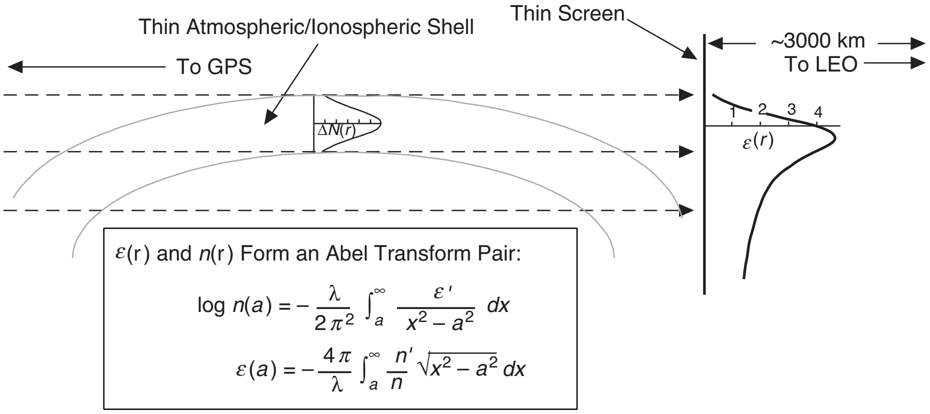
Thin Atmospheric/Ionospheric Shell
Thin Screen
To GPS
~3000 km
To LEO
$\Delta N(r)$
1
2
3
4
$\varepsilon(r)$
ε(r) and n(r) Form an Abel Transform Pair:
$\log n(a) = -\frac{\lambda}{2\pi^2}\int_a^\infty \frac{\varepsilon'}{x^2 - a^2}\,dx$
$\varepsilon(a) = -\frac{4\pi}{\lambda}\int_a^\infty \frac{n'}{n}\sqrt{x^2 - a^2}\,dx$

Fig. 2-31. The relationship between refractivity and the thin-screen phase for a spherical shell.

Fresnel recovery of the electron density profile around the transient. Following the discussion around Eq. (2.10-6) in which the constraint $A(u) \equiv 1$ is formally imposed on the linearized version of the forward Fresnel transform, we first convert it into a discrete form and then invert it using a least-squares algorithm. In this case, one obtains a formal solution for the least-squares estimate of the offsets in the thin-screen phase relative to the nominal profile, which is given by

$$\delta\hat{\varepsilon} = \Gamma \cdot \mathbf{L} \tag{2.10-15}$$

where $\delta\hat{\varepsilon}$ is the $m \times 1$ matrix giving the least-squares estimate of the thin-screen phase offset with $m \le \mathrm{NMAX}/2$. Γ is the $m \times m$ covariance matrix, and $\mathbf{L}$ is the $m \times 1$ information matrix, which for the simple discrete version used here and assuming a stationary white noise process for the observation set (stationarity manifestly does not hold near the transient!), is given by

$$\left.\begin{aligned} &\mathbf{L}_j = \frac{A}{2}\sum_{k=1}^{\mathrm{NMAX}} \beta_k \begin{Bmatrix} \cos\Omega_{jk} + \sin\Omega_{jk} \\ \cos\Omega_{jk} - \sin\Omega_{jk} \end{Bmatrix}^{\mathrm{T}} \begin{Bmatrix} \delta E_k \\ E_k \delta\psi_k \end{Bmatrix} \\ &\Omega_{jk} = \psi_k - \varepsilon_j - B(j-k)^2, \\ &\quad \beta_k = 1, \ \forall k \ne 1 \ \text{ or NMAX}; \ \ \beta_1 = \beta_{\mathrm{NMAX}} = 1/2 \\ &\quad A = \dot{r}_b \Delta T \left(\frac{2}{\lambda D}\right)^{1/2}, \ B = \frac{\pi}{2}A^2, \end{aligned}\right\} \tag{2.10-16}$$

Here NMAX is the number of discrete pairs of amplitude and phase observations selected for the inversion process and NMAX should be at least as great as the Nyquist limit for the selected number of thin-screen points m (but significantly less than the Fresnel aliasing limit $2M^*$). For large NMAX and for a stationary white noise process, $\mathbf{\Gamma}^{-1}$ asymptotically approaches the particularly simple form

$$\left.\begin{aligned} &[\mathbf{\Gamma}^{-1}]_{ij} \to \frac{A^2}{4}\frac{\sin[\varepsilon_i - \varepsilon_j + B(i-j)(i+j-1)]}{\sin[B(i-j)]}, \ i \ne j \\ &[\mathbf{\Gamma}^{-1}]_{ii} \to \frac{A^2}{2}\mathrm{NMAX} \end{aligned}\right\} \tag{2.10-17}$$

which is nearly diagonal when $1 << \mathrm{NMAX} < 2M^*$. A simple trapezoidal rule has been used to convert the continuous forms to discrete forms used in Eqs. (2.10-16) and (2.10-17). This should suffice provided the spatial sampling interval is small (it is about 1/11 of a Fresnel scale for this occultation) and provided the sampled observation vectors are low-pass filtered to eliminate the possibility of certain high-frequency components in their continuous power spectrum leaking in through aliasing. These discrete linearized systems can be applied iteratively using any one of a variety of estimation algorithms (e.g., Householder orthogonal transformation, singular-value decomposition, square-root information filter) until convergence is obtained.

When the converged profile for the thin-screen phase is passed through Eq. (2.10-14), one recovers the refractivity with the a priori constraint $A(u) \equiv 1$ imposed. A comparison of electron density profiles [using Eq. (2.8-1)] with and without this constraint is shown in Fig. 2-32. Steeper gradients in electron density near the transient and a higher peak are significant features of the constrained case. These result in greater defocusing in the constrained case than in the unconstrained case to achieve the observed SNR fadeout and recovery instead of allowing $A(u)$ also to vary to achieve the same effect.

The question of the fidelity of the different Fresnel versions presented here for recovering electron density is moot without first performing a number of simulations with realistic error sources including along-track inhomogeneity and sampling effects. Chapters 3 and 5 of this monograph deal indirectly with the question of the fidelity of the thin-screen model by studying scattering effects based on Maxwell's equations applied to a spherical surface. Fresnel transform theory and the thin-screen model appear to work well in thin atmosphere conditions, which makes certain narrow features observed in the ionosphere good candidates for this approach. Also, tropopause features should be well modeled. But lower troposphere features often violate the thin atmosphere condition for the thin screen, $1 + a\alpha(d\alpha / da) > 0$ or $\beta < (2\pi)^{-1/2} \approx 0.4$, and they even exceed the super-refractivity threshold $\beta = 1$ on occasion. The impact parameter space curve provides a better platform for these lower troposphere features, working exceptionally well except for super-refractive layers.

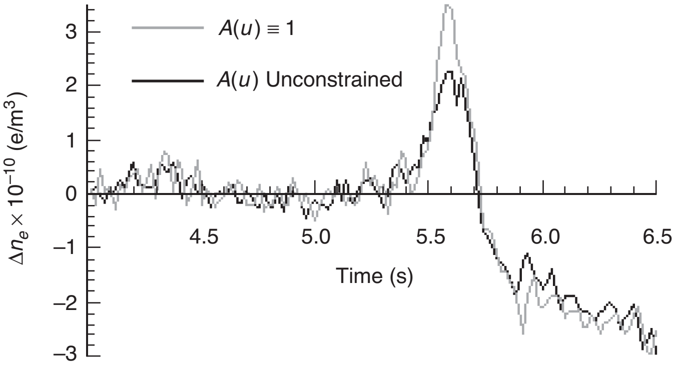

Fig. 2-32. Fresnel recovery of the transient in electron density distribution for the occultation shown in Fig. 2-15 around $t = 5.6$ s, with and without adjoining the a priori constraint $A(u) \equiv 1$.

For observation epochs greater than $t = 6.5\,\mathrm{s}$ shown in Fig. 2-32, the recovered profile for Δn_e gradually levels off as the tangency point of the ray descends into the upper mesosphere and into a more neutral medium. This regime where Δn_e levels off could provide a zero point from which the absolute value of the electron density near the transient can be estimated. However, variability from the ionosphere at large may corrupt this estimate. Figure 2-14 provides one example of the trend in bending angles for the upper mesosphere.

References

[1] G. Lindal, J. Lyons, D. Sweetnam, V. Eshleman, D. Hinson, and G. Tyler, "The Atmosphere of Uranus: Results of Radio Occultation Measurements with Voyager 2," *Journal of Geophysical Research,* vol. 92, pp. 14,987–15,001, 1987.

[2] G. Lindal, "The Atmosphere of Neptune: An Analysis of Radio Occultation Data Acquired by Voyager 2," *The Astronomical Journal,* vol. 103, pp. 967–982, 1991.

[3] M. Born and E. Wolf, *Principles of Optics,* 6th ed., Oxford, United Kingdom: Pergamon Press, 1980.

[4] W. G. Melbourne, E. S. Davis, C. B. Duncan, G. A. Hajj, K. R. Hardy, E. R. Kursinski, T. K. Meehan, L. E. Young, and T. P. Yunck, *The Application of Spaceborne GPS to Atmospheric Limb Sounding and Global Change Monitoring*, JPL Publication 94-18, Jet Propulsion Laboratory, Pasadena, California, April 1994.

[5] E. Marouf, G. Tyler, and P. Rosen, "Profiling Saturn's Rings by Radio Occultation," Icarus, vol. 68, pp. 120–166, 1986.

[6] E. Bramley, "The Diffraction of Waves by an Irregular Medium," *Proceedings of the Royal Society*, vol. A225, pp. 515–518, 1954.

[7] E. Saltpeter, "Interplanetary Scintillations. I. Theory," *The Astrophysical Journal*, vol. 147, pp. 433–448, 1967.

[8] E. N. Bramley, "The Accuracy of Computing Ionospheric Radiowave Scintillation by the Thin-Phase Screen Approximation," *Journal of Atmospheric and Solar-Terrestrial Physic*s, vol. 39, pp. 367–373, 1977.

[9] W. Hubbard, J. Jokopii, and B. Wilking, "Stellar Occultations by Turbulent Planetary Atmospheres: A Wave-Optical Theory Including a Finite Scale Height," *Icarus*, vol. 34, pp. 374–395, 1978.

[10] D. Knepp, "Multiple Phase-Screen Calculation of the Temporal Behavior of Stochastic Waves," *Proceedings of the IEEE*, vol. 71, no. 6, pp. 722–737, 1983.

[11] J. Martin and S. Flatté, "Intensity Images and Statistics from Numerical Simulation of Wave Propagation in 3-D Random Media," *Applied Optics*, vol. 27, no. 11, pp. 2111–2126, 1988.

[12] J. Goodman, *Introduction to Fourier Optics*, New York: McGraw-Hill, 1968.

[13] A. J. Devaney, "A Filtered Backpropagation Algorithm for Diffraction Tomography," *Ultrasonic Imaging*, vol. 4, pp. 336–350, 1982.

[14] E. Karayel and D. Hinson, "Sub-Fresnel Scale Vertical Resolution in Atmospheric Profiles from Radio Occultation," *Radio Science*, vol. 32, no. 2, pp. 411–423, 1997.

[15] M. E. Gorbunov, A. S. Gurvich, and L. Bengtsson, *Advanced Algorithms of Inversion of GPS/MET Satellite Data and Their Application to Reconstruction of Temperature and Humidity*, Report 211, Max Planck Institute for Meteorology, Hamburg, Germany, 1996.

[16] G. Watson, "The Diffraction of Electric Waves by the Earth," *Proceedings of the Royal Society*, vol. A95, pp. 83–99, London, United Kingdom, 1918.

[17] J. Jackson, *Classical Electrodynamics,* 2nd ed., New York: John Wiley & Sons, Inc., 1975.

[18] M. Levy, *Parabolic Equation Methods for Electromagnetic Wave Propagation*, London, United Kingdom: The Institution for Electrical Engineers, 2000.

[19] G. A. Hajj, E. R. Kursinski, W. I. Bertiger, S. S. Leroy, T. Meehan, L. J. Romans, and J. T. Schofield, "Initial Results of GPS–LEO Occultation Measurements of Earth's Atmosphere Obtained with the GPS/MET Experiment," *Proceedings of the IAG Symposium*, G1, GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications, IUGG XXI General Assembly, Boulder, Colorado, July 2–14, 1995, Springer-Verlay, pp. 144–154, 1996.

[20] M. Gu and F. Brunner, "Theory of the Two Frequency Dispersive Range Correction," *Manuscripta Geodœtica*, vol. 15, pp. 357–361, 1990.

[21] S. Bassiri and G. Hajj, "Higher-Order Ionospheric Effects on the Global Positioning System Observables and Means of Modeling Them," *Manuscripta Geodœtica*, vol. 18, pp. 280–289, 1993.

[22] V. V. Vorob'ev and T. G. Krasil'nikova, "Estimation of the Accuracy of the Atmospheric Refractive Index Recovery from Doppler Shift Measurements at Frequencies Used in the NAVSTAR System," *Physics of the Atmosphere and Ocean*, vol. 29, no. 5, pp. 602–609, 1994.

[23] T. Meehan, personal communication, Jet Propulsion Laboratory, Pasadena, California, 1995.

[24] W. Melbourne, *Sensing Atmospheric and Ionospheric Boundaries in GPS Radio Occulatation Observations From a Low Earth Orbiter, Part 1*, JPL Publication 96-5, Jet Propulsion Laboratory, Pasadena, California, July 1998.